Eötvös Loránd University of Sciences

Doctoral School of Biology

Director: Dr. Anna Erdei, MHAS

Evolution biology and Theoretical Biology Doctoral Program

Director: Dr. Eörs Szathmáry, CHAS

Department of Plant Systematics, Ecology and Theoretical Biology

Director: Dr. János Podani, CHAS

Anna Fedor

Linguistic recursion

Learning of recursion in artificial grammars by humans and a neural network

PhD dissertation

Supervisor:

Dr. Eörs Szathmáry, professor, CHAS

2012

1 Contents

2 Introduction

2.1 Recursion in general

In this subchapter, I summarize what recursion broadly means at different areas of life mainly based on Corballis's book, *The recursive mind* (Corballis, 2011). Other, more specific meanings, including linguistic recursion is covered in the next subchapter.

2.1.1 What is recursion?

A google search for recursion suggests: "Did you mean: *recursion*". This geek joke is a variant of a funny dictionary entry "Recursion: see *Recursion*", which now appears in many online dictionaries (e.g., urbandictionary.com) and even textbooks (e.g., Kernighan & Ritchie, 1988, p. 269). These "definitions" capture the idea of the infinite loop inherent in recursion.

Visual representations of recursion (Figure 1) include fractals, for example a tree-like structure where the stem at the bottom branches out to two stems, each of which branches out to two stems, each of which branches out…, etc. A more everyday example is the box of Droste cocoa powder (http://en.wikipedia.org/wiki/Recursion), on which a nurse is holding a tray with a box of Droste cocoa powder on which a nurse is holding a tray with a Droste cocoa powder… ad infinitum. The funniest (though not infinite) example I have seen is a photograph about an escalator with a sign above with a pictogram showing a man hit his head in the sign above the escalator. If there was no sign, there wouldn't be need for a sign for warning people that there is a sign.

Talking about these examples invokes recursive sentence structures too. It seems like repeating the same thing over and over again, but with each repetition, the story becomes more and more complex (Corballis, 2011). In the case of the cocoa powder box, we not only repeat "on which a nurse is holding a tray with a box of Droste cocoa powder" but also understand that the box held by the first nurse is the same box on which the second nurse is depicted. It is our understanding of the deepening loop that makes these sentences more complex than simple repetition and what captures our imagination.

It might be clear from these examples that a main characteristic of recursion is that it takes its own output as the next input. In the fractal example on Figure 1a, the input is a stem, the outputs are two branches, both of which can be taken as the input of the next branching procedure. The other main characteristic is that in theory, the recursive procedure can be repeated infinitely.

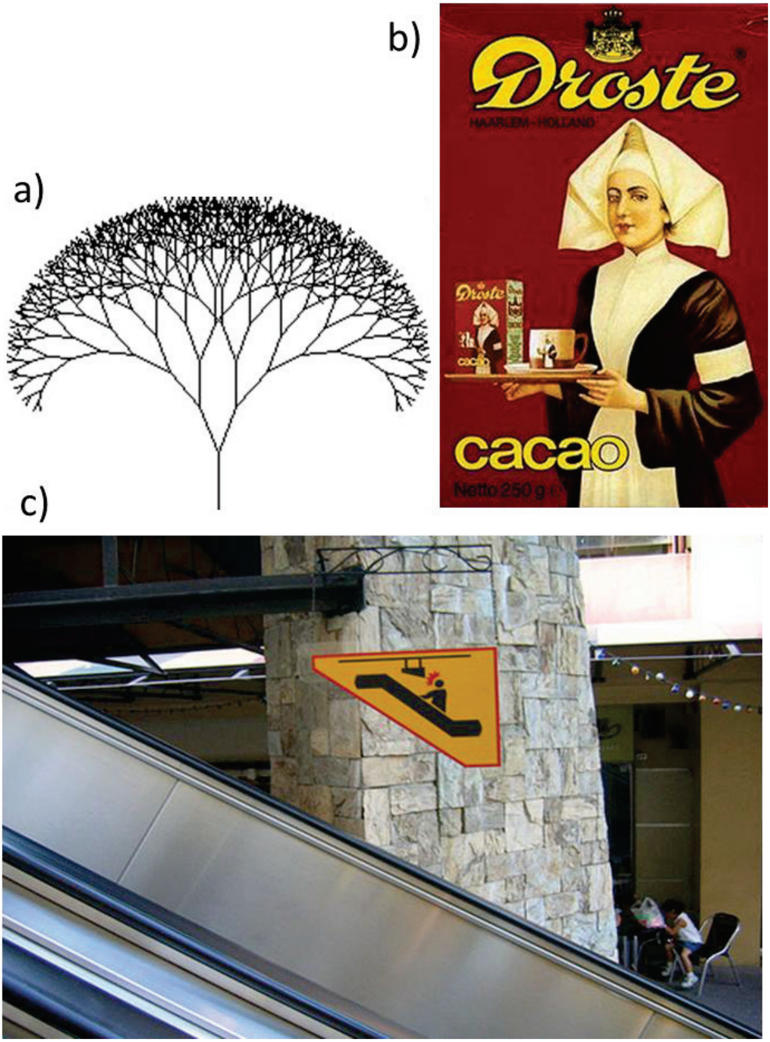


Figure 1. Examples of visual representations of recursion. a) A fractal resembling a tree; b) The Droste cocoa-powder box; c) A sign hanging above an escalator.

2.1.2 Why is recursion interesting?

We always wanted to know what makes us special, superior to other animals. It is clearly not our physique; there are bigger, stronger, faster animals on Earth. It is rather our minds that enabled us to populate the globe and use its resources to our advantage. But what is it about our mind that is special? What makes us human? Among other properties, like theory of mind, self-awareness, tool-making, ownership, music, etc., language is undoubtedly something that is uniquely human. Language makes it possible to cooperate in complex situations[1], to plan the future and to pass on our culture from generation to generation and thus accumulate knowledge about the world.

Animals communicate too, but their ways of communication is much simpler than human language. Human language can be characterized as the critical combination of symbolic reference with complex syntax (Szathmáry et al., 2007). Symbolic reference means that meanings are associated with symbols (words, phrases, signs, etc.; this is the focus of semantics) and syntax is the set of grammatical rules for combining symbols into phrases and phrases into sentences. The question is: do animals have either semantics or syntax?

Vervet monkeys might be the closest to naturally have something similar to words with meanings: they have different alarm calls for snakes, leopards and eagles (Cheney & Seyfarth, 1990). Language-trained apes (e.g., Gardner & Gardner, 1969; Patterson, 1978), dolphins (Herman, Richards, & Wolz, 1984) and a grey parrot (Pepperberg, 1990) are reported to be able to learn few hundred words or signs, although unlike human children, they need extensive training. This vocabulary is orders of magnitude smaller than that of an average adult, but nevertheless would be enough to compose sentences with complex syntax. However, the syntax of these animals is on the level of protolanguage at best. Protolanguage includes the simple combination of symbols without real grammar (D. Bickerton, 1995) which is characteristic of children below two, pidgin language speakers and the language of those individuals who were deprived of language during the sensitive period, but were taught language later (Kaspar-Hauser syndrome; Derek Bickerton, 1990). Sentences consist of 2-3 words, often repeated many times in random order, and grammatical words are almost non-existent.

---

[1] According to (Derek Bickerton, 2009; Derek Bickerton & Szathmáry, 2011) language and cooperation had a common evolutionary triggering episode, namely confrontational or high-level scavenging.

Even if we accept that language trained animals are capable of simple syntax, this is very different from what we find in human languages. Unlike animal communication, human language is open-ended. It has the feature that is called discrete infinity: we can recombine meaningful discreet units (morphemes and words) into an infinite number of larger structures (phrases and sentences). Most linguists agree that the key to this open-endedness is recursion, because it makes it possible to embed phrases within phrases infinitely in principle. As such it is of utmost importance for the evolution of language: it could be the ingredient that provided the means for the last major transition in evolution from animal communication and groups to human language and society (Fedor, Ittzés, & Szathmáry, 2009; Maynard Smith & Szathmáry, 1995).

An example for increasing complexity through multiple embeddings could be the following series of sentences:

(1) I know that your birthday party is today. (Of course I know, since you made me put it in my diary.)
(2) You know that I know that your birthday party is today. (Because I texted you today to ask what to bring.)
(3) I know that you know that I know that your birthday party is today. (Because you replied my text message and asked me to bring a bottle of wine.)

This could go on forever. Although each sentence has a finite number of words, we can always compose a longer sentence by adding another "I know" or "You know" at the beginning. Of course it will be a little hard to follow after a while due to both practical limits set by our working memory and the increasing complexity of the sentences which is difficult to understand.

Language needs to be recursive to be able to express our recursive thoughts (Steven Pinker & Jackendoff, 2005). Vice versa, without being able to think recursively, we would not understand the increasing complexity of the above sentences. Recursive thinking in turn is necessary for theory of mind and mental time travel among others (Corballis, 2011). Theory of mind is the ability to impute mental states (beliefs, intents, desires, pretending, knowledge, etc.) to oneself and others (Premack & Woodruff, 1978). Zero order theory of mind includes the awareness of one's own mental states, while first order theory of mind refers to knowing or thinking what others know or think and therefore is recursive (Corballis, 2007c). While there is a debate about whether chimpanzees have zero and first order theory of mind, most researchers agree that no animal species have higher order theory of mind that is necessary for example, to understand sentence (3). Mental time travel is the ability to call past episodes to mind (episodic memory) and to imagine future episodes. According to Corballis, mental time travel can be a recursive operation in that imagined episodes can be inserted into present consciousness, and imagined episodes can even be inserted into other imagined episodes (Corballis, 2011). Mental time travel allows us to learn from past events and to plan for the future. It is also responsible for foreseeing the future and as a consequence for the awareness of our own death. As such it might have been one of the factors leading to religious believes about life after death.

According to Corballis (2007c), recursive thinking is the primary characteristic that distinguishes the human mind from that of other animals. It not only makes recursive linguistic rules possible but also enables us to travel mentally in time and think about what others think. He also assumes that recursion is necessary for counting or tool-making. However, his examples reveal a quite broadly conceived definition of recursion: something within something similar. Other, more specific definitions of recursion also exist and will be explored in the next section.

2.2 The many meanings of recursion

According to Lobina (2011a, 2011b)[2] recursion (broadly meaning self-reference) can be applied to four different constructs:

1. Definition by induction, as in mathematics
2. A feature of real-time processes when these include an operation that calls itself
3. The architectural attribute of structures, an X within an X
4. Recursion, as a general property of computational systems

Fitch (2010) identifies three meanings:

5. Recursion in meta-mathematics
6. Recursion in computer science
7. Recursion in linguistics

[2] (Lobina, 2011b) is an unpublished PhD thesis and page numbers were not available at the time of writing. All quotations are from Chapter 1 and 2.

Meaning 5 deals with recursive function theory; a misleading designation because its subjects (recursive functions) do not involve recursion in the linguistic or computer science senses. The name itself is being replaced by computability theory and I will not deal with this construct here. Recursion in mathematics (meaning 1) and in computer science (meaning 6) is not directly related to this thesis either, but I will briefly summarize them for illustrating the common feature of the different meanings. According to Lobina, recursion as a general property of computational systems (meaning 4) is mostly related to the claim that recursion is responsible for the discrete infinity of language and as such, relevant for the evolution of language. For the purposes of this thesis, Fitch's recursion in linguistics (meaning 7) is the most important, which includes meanings 2 and 3 in Lobina's categorization and will be described in detail in section 2.2.2.

2.2.1 Recursion in mathematics and computer science

In mathematical logic and computer science, a recursive definition (or inductive definition) "consists in defining a function by specifying each of its values in terms of previously defined values" (Cutland, 1980, p. 32, as cited in Lobina, 2011a).

For example, the factorial function *n!* can be defined as:

i. if $n = 1$, then $n! = 1$
ii. if $n > 1$, $n! = n \times (n-1)!$

The first equation is the base case; the second is the recursive step (or inductive clause). Note, that the latter contains the factorial function on both sides of the equation, i.e. self-reference is a defining feature of this construct. The results of self-reference is that "chains of unfinished tasks develop, which automatically yields hierarchy among the operations so produced" (Lobina, 2011b), i.e., for *2!* First *0!* then *1!* then *2!* must be calculated.

However, there is "nothing intrinsically recursive about the factorial"; it is the nature of the solution that is recursive here (Lobina, 2011b). Structural and generative recursion must not be conflated. Of course, recursive implementations are especially well-suited to operate over recursive structures , but memory limitations or efficiency issues of the machine that is implementing the algorithm can make iterative implementations necessary (Lobina, 2011b).

For example, *4!* can be iteratively computed if we keep a running product together with a counter (Lobina, 2011a): multiply the product by the counter, and then increase the counter by 1. See Table 1 for a comparison of computational steps during the recursive and iterative implementations of *4!*. Both recursion and iteration operate over their own outputs, but self-reference is only involved in recursion (Lobina, 2011b).

Table 1. Recursive and iterative implementations of 4!

	Recursion	Iteration		
		Counter	Prev. prod.	Product
Step 1	4 × (3!)	1 ×	1 =	1
Step 2	4 × (3 × (2!))	2 ×	1 =	2
Step 3	4 × (3 × (2 × (1!)))	3 ×	2 =	6
Step 4	4 × (3 × (2 × 1))	4 ×	6 =	24
Step 5	4 × (3 × 2)			
Step 6	4 × 6			
Step 7	24			

According to Fitch (W. Tecumseh Fitch, 2010, p. 76) in computer science a "recursive function is one which calls itself (that is, where a command to run function x appears within the definition of function x itself)". His interpretation agrees with that of Lobina, in that he also acknowledges that recursion cannot be determined by observing the string that was produced. The same-input-output behaviour can be the result of both recursive and non-recursive implementations. Fitch provides a recursive and iterative implementations for defining A^nB^n strings which will be important later. The recursive function is:

```
define function AnBn(n):
if n is 1
        then return "AB"
else
        return ("A" + AnBn(n-1) + "B");
```

And the iterative function is:

```
define function AnBn(n):
integer counter i;
```

```
A_section = "A";
B_section = "B";
if n > 1 then {
        for (i = 2) to (i = n)
                A_section = A_section + "A";
                B_section = B_section + "B";
        end
}
return A_section + B_section
```

These two functions return the same strings for any *n* (e.g., for n = 3, they both return AAABBB) thus it is not possible to tell which function produced a string of the form A^nB^n. However, Fitch assumes that there is an underlying implicit structure to the strings, which in the case of the recursive function is centre-embedded (because AB strings are embedded within other AB strings) and in the case of iteration is a "flat tree" (W. Tecumseh Fitch, 2010, p. 77).

2.2.2 Linguistic recursion

2.2.2.1 Recursion in production rules

In formal language theory rewrite rules represent transformations where strings on the left hand side of the arrows are replaced with strings on the right hand side of the arrows. The example most usually used in recursion-related studies is a sentence that is part of the well-known nursery rhyme *The house that Jack built*[3]:

---

[3] The poem starts with a simple sentence "This is the house that Jack built" and then gradually expands by adding phrases one-by-one. The last verse goes like this:

This is the horse and the hound and the horn
That belonged to the farmer sowing his corn
That kept the cock that crowed in the morn
That waked the priest all shaven and shorn
That married the man all tattered and torn
That kissed the maiden all forlorn
That milked the cow with the crumpled horn
That tossed the dog that worried the cat
That killed the rat that ate the malt

(4) This is the cat that killed the rat that ate the malt.

This sentence can be produced by the following rewrite rules:

i. $S \rightarrow N\ VP$

ii. $VP \rightarrow V\ NP$

iii. $NP \rightarrow N\ (CP)$

iv. $CP \rightarrow C\ VP$

v. V → *is, killed, ate*

vi. N → *this, the cat, the rat, the malt*

vii. C → *that*

where S = sentence, NP = noun phrase, VP = Verb Phrase, V = verb, N = noun, CP = complementer phrase, C = complementer. All of these are non-terminals that represent phrase types or word classes. Actual words that can appear in sentences and cannot be replaced further (i.e. they never appear on the left hand side of the arrows) are called terminals and are in italics. Figure 2 depicts the parse tree of the above sentence: it shows how the sentence was derived by applying the rules of the grammar and what the structure of the generated sentence is.

Rules ii-iv are indirectly recursive, because they can be applied infinitely in cycles. Direct recursion is also possible, for example $NP \rightarrow N\ (NP)$, which would generate embeddings like *I saw Vax's [girlfriend's [mother's glasses]]*. This fulfils the mathematical definition of recursion, i.e., NP appears on both sides of the arrow, just as in the example above about the factorial.

Recursion can be categorized based on the "symmetry" of the embedding to tail-recursion (or end recursion) and centre-embedded recursion (CER). In tail-recursion, new phrases are inserted at the end or beginning of the previously inserted phrase, as in sentence (4), while in CER phrases are inserted within previously inserted phrases:

(5) This is the malt the rat the cat killed ate.

Figure 3 represents the structure of this sentence. With the addition of one more production rule, it can be produced by the same rewrite rules as sentence (4). This rule is responsible for the centre-embedding:

That lay in the house that Jack built!

viii. CP → NP VP

CER seems to be present in all human languages, but not in animal calls, and it also separates finite state and phrase structure grammars described in the next section.

2.2.2.2 Chomsky hierarchy

Formal grammars (the collection of production rules[4]) and languages (the collection of strings generated by these grammars) can be categorized based on the type of production rules they employ (Chomsky, 1956). Rules that can take any form (any terminal or nonterminal symbol can appear on both sides of the arrows) are non-restricted, while other rules might conform to some restrictions on their forms. These restrictions in turn influence the expressive power of the grammars (i.e. the complexity they can describe). The hierarchical categorization of formal or generative grammars based on the restrictions on the form of their production rules is called the Chomsky hierarchy (Table 2).

If all the rules of a grammar contain only a nonterminal symbol on the left hand side, it means that each of these rules can be applied anywhere in the strings (sequences of symbols), independently of the symbols around the nonterminal. These grammars are called context-free grammars (CFG). If the left hand side contains more than one symbol, then the grammar is context-sensitive (CSG). The rules of finite state grammars (FSG) can take only three possible forms: the left side is a single nonterminal symbol, while the right side may be the empty string, a single terminal symbol, or a single terminal symbol followed by a nonterminal symbol. Finally, there are no restrictions on the forms of the rules of unrestricted grammars. The containment hierarchy of these grammars is: unrestricted grammars $\supset$ CSG $\supset$ CFG $\supset$ FSG.

[4] The rules describe how to form strings from the alphabet of the language that are valid according to the language's grammar. A grammar does not describe the meaning of the strings, only their form. Generative grammars along with analytic grammars are subcategories of formal grammars.

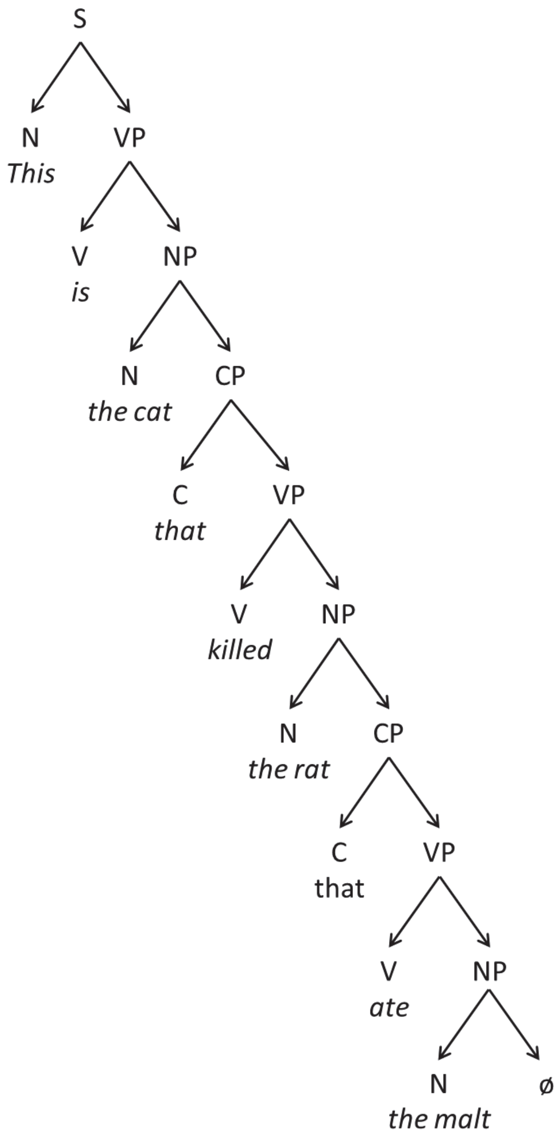


Figure 2. The parse tree of sentence (4). The analysis could be more precise with further dividing "the cat" and other similar phrases to determiners and nouns "the" and "cat", respectively.

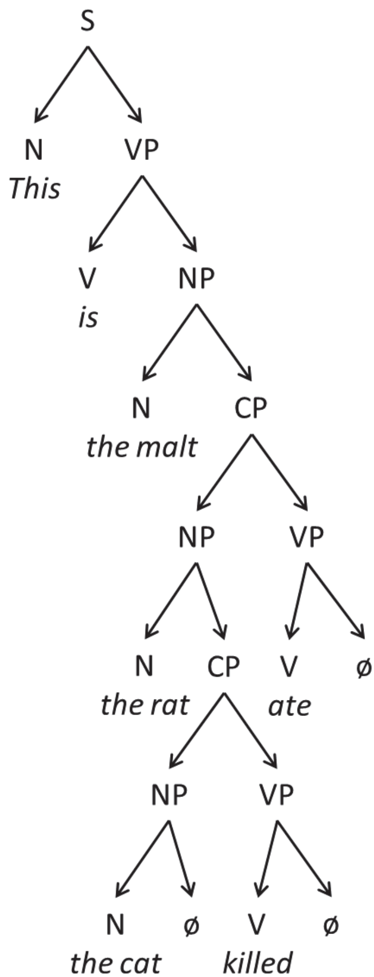


Figure 3. The parse tree of sentence 5.

Table 2. The Chomsky hierarchy. Upper-case and lower-case symbols stand for nonterminals and terminals, respectively; ε is an empty string; α, β, γ are arbitrary strings of terminal and non-terminal symbols (Jurafsky & Martin, 2007).

Type of grammar	Name of grammar	Name of language	Constraints on production rules
Type 0	unrestricted grammars	recursively enumerable languages	$\alpha \rightarrow \beta; \alpha \neq \varepsilon$
Type 1	context-sensitive grammars (CSG)	context-sensitive languages	$\alpha A\beta \rightarrow \alpha\gamma\beta; \gamma \neq \varepsilon$
Type 2	context-free grammars (CFG)	context-free languages	$A \rightarrow \gamma$
Type 3	finite state grammars (FSG)	regular languages	$A \rightarrow aB$ or $A \rightarrow a$

All grammars above FSG level are called phrase structure grammars (PSG) and all of these are able to generate string with an underlying recursively-embedded, hierarchical structure. “There is a broad consensus in linguistics and machine learning both that PSGs are more powerful than FSGs, and that grammars at the context-free level are, minimally, a crucial component of all human languages” (W. Tecumseh Fitch & Hauser, 2004, supporting online material). FSGs are limited to local dependencies between words and are therefore inadequate to describe the long-range dependencies in CER. For example, in sentence (5) the words *rat* and *ate* are separated by another phrase that was inserted between them (*the cat killed*), whereas in sentence (4) dependencies stay local and are not separated by other phrases (Figure 4). Human languages are at the PSG level and include the capacity for recursive embedding of phrases within phrases thus creating long-range, hierarchical relationships (Hauser, Chomsky, & Fitch, 2002).

According to Fitch (W. Tecumseh Fitch, 2010), a hypothetical protolanguage equipped with FSG level syntax that combined nouns and verbs and the ability to gain open-ended shared vocabulary would allow its users to express a wide range of concepts. However, it would be stuck at a “pregiven canonical level of specificity or generality” (p. 89). Recursive embedding of phrases within phrases provides flexible and unbounded means of expressing any thoughts with arbitrary degrees of accuracy and abstraction.

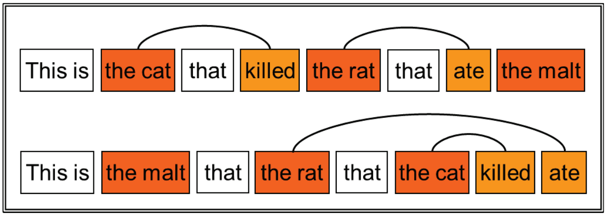


Figure 4. Top: Local dependencies in a sentence with tail-recursion; Bottom: one local and one long-range dependency in a sentence with CER.

2.2.2.3 Recursion as a property of structures

According to Pinker and Jackendoff (2005, p. 203) "recursion refers to a procedure that calls itself, or to a constituent that contains a constituent of the same kind". This definition refers to two different meanings. The first one (a procedure that calls itself) can be applied to production rules: a production rule (and thus a grammar) is recursive if it contains the same symbol on both sides of the arrow. The second refers to structures: more generally, a recursive structure is something that contains something of the same kind. These structures are also called self-embedded, which is a special case of nested (or embedded) structures. In nested structures one constituent falls totally within another constituent, and they are not always the same kind. Nested structures are self-embedded structures if the two constituents are the same kind (Lobina, 2011b). According to the above definition, self-embedded structures are recursive, while nested structures are not.[5]

According to the first half of the definition, certain rewrite rules are recursive, as it was described above. One would think that sentences generated by recursive rules are in turn recursive themselves. However, formal grammars model *weak generative capacity* (i.e., the generation of strings abstracted away from the underlying structure of sentences), as opposed to *strong generative capacity* (i.e., the generation of structure; Lobina, 2011b). In other words, there is no way to tell, whether a string was produced recursively or iteratively. Nor it is possible to automatically conclude that successful

[5] For Pinker and Jackendoff "same kind" refers to the category of the element that heads a constituent, however, at a higher level of abstraction every syntactic phrase is a self-embedded structure as in [Specifier [Head – Complements]] (Lobina, 2011b). In this sense, recursion is the general property of all human languages, independently whether they contain self-embedded structures in the stricter sense or not.

parsing of a string that was generated recursively means that the parser applied recursion too. Because of these distinctions, it is better not to automatically use the word *recursive* to refer to strings that were generated or parsed by recursive procedures.

Structures can be recursive in their own way. A structure is recursive if it "includes an abstraction of itself (an X within an X)" (Lobina, 2011a, p. 3) – a definition that when applied to linguistic constituents gives the second half of Pinker and Jackendoff's definition (a constituent that contains a constituent of the same kind). Parse trees underlying strings can be recursive, e.g., the tree on Figure 3 is recursive because a VP contains an NP that contains a CP which in turn contains a VP that contains an NP that contains a CP again. The recursive nature of this structure reflects the recursive nature of the rules that were used to generate it. So it seems that recursive rules generate recursive structures. The question is: do strings generated by recursive production rules have an underlying recursive structure? Can strings be recursive at all? Let's examine these questions in the next section.

2.2.2.4 Recursion in artificial grammar learning tasks

Artificial grammar learning (AGL) tasks are widely used to test the abilities of different species in learning different grammatical rules. During these tasks participants are usually trained and tested on a set of artificial sentences to assess whether they could master the grammatical rule underlying these sentences. Sentences are composed of a set of nonsense "words" (the vocabulary), which could be anything from actual letters to geometrical shapes but are usually consonant-vowel (CV) syllables. The theory behind this paradigm is that removing semantics from a language makes it possible to research its pure syntax. One of the advantages of this method that humans and animals can be tested with basically the same stimuli and their performances could be compared (if we make sure that the words used are easily distinguishable to both species).

The focus of this thesis is CER so let's take a simple centre-embedded sentence to illustrate how this works:

(6) The rat that the cat killed squeaked.

Phrases here consist of noun-verb pairs meaning who did what*: cat-killed*, *rat-squeaked*. These word-pairs are connected by within-phrase dependencies which can be either short-range (as in *cat-killed*) or long-range (as in *rat-squeaked*). There is also a between-phrase dependency, namely that establishes *the rat that squeaked* is the same rat that *the cat killed*, i.e., *rat* is the object of the embedded phrase. To understand what happened, i.e., who did what to whom one has to understand the centre-embedded (recursive) structure of this sentence at least implicitly. In sum, to correctly parse this sentence, recognizing CER is necessary, but it does not mean that parsing as a process is recursive itself.

How can this kind of centre-embedded structure be modelled by artificial languages? The problem is that sentences taken from natural languages have an underlying structure, which, in this case, involves self-embedding, while rewriting rules does not return structures at all, only strings. The hierarchical tree structure underlying recursion can reflect the way the structure was produced or the way the structure is supposed to be parsed; but recursion is not the property of the structure itself and definitely not the property of strings (Lobina, 2011b). In natural languages morphology (e.g., agreement between singular and plural nouns and verbs) and semantics (e.g., rats squeak not cats) ensure that sentences are parsed with the intended structure. Is there anything in artificial languages that does the same?

In the first generation of AGL experiments on recursion after the influential paper of Hauser, Chomsky and Fitch (Hauser, et al., 2002) nothing ensured within and between phrase dependencies. Centre-embedded sentences followed the formula of A^nB^n (e.g., *AABB* when n = 2), where *A* and *B* were two distinct class of artificial words (modelling nouns and verbs in this case). These sentences could have been generated by the following pair of CFG rules:

i. $S \rightarrow aSb$
ii. $S \rightarrow ab$

The rule is recursive and it was assumed that the structure of the strings is recursive too thus successful grammaticality judgement proves recursive parsing. However, as critiques pointed out, these strings could be parsed by simply counting *A*s and *B*s and matching the two numbers. For this reason these type of strings were named *counting recursion* as opposed to real CER.

Corballis (2007a, p. 1582) suggested using cues to make sure that the parser connects A-B pairs:

> To distinguish between these models, information beyond the strings themselves is required. That information might be semantic, or prosodic, or

> perhaps even some neurophysiological process that makes embedding mandatory. It might also depend on associative learning. For example, A^nB^n strings might be parsed as centre-embedded if there were established associations between AB pairs from the outside in—such that there is an association between the first and last, second and second to last, and so on.

The second generation of AGL experiments tried to ensure the connection by statistical co-occurrence (each *A* word has one or few *B* pairs and vice versa) usually complemented by phonological cues. In these sentences, the underlying centre-embedded structure must be understood for successful parsing. Most scholars (e.g., Corballis, 2007b; W. Tecumseh Fitch, 2010), but not Lobina (Lobina, 2011a, 2011b) believe that with this modification one can make sure that strings are parsed using the underlying structure that was intended by the experimenter.

There is one more component of natural sentences that could be modelled to bring artificial grammar models of CER closer to natural language. It is the between phrase dependencies that are the result of the calling function of recursion from one phrase to the next. It is hard to imagine how it would be possible to do this without bringing meaning to the picture. This would be the task of the third generation of AGL experiments.

2.2.2.5 Recursion and iteration

I mentioned at the end of section 2.2.2.1 that there are two types of recursion: tail-recursion and CER. The definition for procedures is the following (Steven Pinker & Jackendoff, 2005, p. 203): “in tail-recursion, a procedure invokes another instance of itself as a final step”, while in CER “a procedure invokes an instance of itself in mid-computation and then must resume the original procedure from where it left off”. For linguistic structures the definition translates to the following: in tail-recursion the location of the embedded constituent is at the beginning or at the end of the embedding constituent; while in CER, the embedding happens in the inside of the embedding constituent (Steven Pinker & Jackendoff, 2005).

Some authors confound tail-recursion with iteration, saying that tail-recursion is not even true recursion. This confusion is the consequence of confounding procedures, structures and strings. In natural language tail-recursion is very easily distinguishable from iteration. In tail-recursion, just like in CER there is a calling function between phrases: a person correctly parsing the sentence below understands that it is that particular dog that worried exactly that cat that killed that rat that ate the malt.

(7) This is the dog that worried the cat that killed the rat that ate the malt that lay in the house that Jack built.

An iterative version of this sentence would be:

(8) A dog that worried a cat, a cat that killed a rat, a rat that ate some malt, some malt that lay in the house, a house that Jack built.

Each phrase takes the next one as a reference but at the next step, the first phrase is forgotten (Corballis, 2011). In tail-recursion, there is dependency between phrases; the phrase to be built in next might be called by the previous phrase (Corballis, pers. comm.) – this dependency is missing in iteration. The common feature of iteration and recursion is that infinite structures can be achieved by both.

When recursive sentences are modelled by artificial sentences, the dependency between phrases gets lost, just as in the case of CER. Strings generated by a tail-recursive procedure can be interpreted by the parser as simple iteration. As Pinker and Jackendoff (2005, p. 203) put it: "Tail-recursion can be mimicked (at least in input–output behaviour or 'weak generative capacity') by a computational device that implements simple iteration, where one instance of a procedure can be completed and forgotten by the time the next instance has begun. Tail-recursion, however, cannot be mimicked by iteration when it comes to computations that require more than duplicating input–output behaviour ('strong generative capacity'), such as inferences that depend on the grouping and labelling of constituents."

Repetition is an even simpler procedure with which infinite strings can be generated, but it does not make the signal more complex. E.g., in the sentence "Jack built a very very very big house" the repetition of *very* does not add extra information.[6] Repetition is also present in animal calls, like birdsongs or the pant-hoots of chimpanzees. The function of repetition can be emphasising, or simply prolonging the signal (Corballis, 2011).

[6] Of course there are languages where repetition does have a meaning. E.g., in most Indian languages repetition has various functions and modifies meaning (Sankaranarayanan, 2002); whereas in Indonesian repetition of nouns indicates plural.

2.2.3 Recursion, as a general property of computational systems

According to Lobina, there is one more meaning of recursion in linguistics: the global recursive property of language as a production system which is not the same as the recursive property of certain rewrite rules. "[the recursive property of rewrite rules] is an internal application within production systems, which ought to be kept distinct from the global recursive property of collections of rewriting rules qua production systems" (Lobina, 2011b). It is recursion in this global meaning that is responsible for the discreet infinity of human languages in general. It is true, that recursive rewrite rules can generate infinite number of sentences, but this can be achieved with other, non-nested structures too, like coordinative sentences (Lobina, 2011b); e.g., *The dog worried the cat and the cat killed the rat and the rat ate the malt*. Recursion in the global sense allows for the discrete infinity of language; thus studies that examine nested structures in relation to unboundedness are mistaken.

After reviewing Chomsky's numerous works, Lobina (Lobina, 2011b) concluded that for Chomsky, recursion is simply a mapping between sound and meaning and as such refers to the mechanism of sentence production, not the structure of the sentences. For example, in Chomsky, 1995 (p. 226, as cited in Lobina, 2011b) he defined Merge as a procedure that "recursively constructs syntactic objects from [lexical] items" and other syntactic objects and claimed that recursive Merge is the only fixed computational procedure that underlies all languages, i.e., it is a recursor. This recursive property of Merge is "entirely independent of the character of the structures it generates" (Lobina, 2011b). It seems that for Chomsky recursion only "means the need to enumerate the potentially infinite number of expressions" (Lobina, 2011b).

It seems to me that this distinction is not very clear to most scholars (including myself). However, it makes sense if we follow the reasoning below:

i. PSG rewrite rules can be recursive while FSG rewrite rules are not.
ii. Human languages are on the PSG level; FSGs are not able to describe them because of the lack of recursive rewrite rules.
iii. Human languages are unbounded.
iv. If sentences can be of arbitrary length, FSGs can generate infinite number of sentences too (just as PSGs), so finite state languages are unbounded too in this sense.

v. So recursion that is responsible for unboundedness is a different kind of recursion than the one that is responsible for the FSG-PSG distinction.

Whether Chomsky meant two different kinds of recursion or not there is another solution to this conundrum. It can be unboundedness or discrete infinity that has two different meanings. One meaning is that the number of possible sentences is infinite. It follows simply from the fact that sentences are of arbitrary length and compositional[7] (we can always compose a longer sentence by adding a phrase either recursively or iteratively). Animal calls can be infinite in this sense in theory. The other meaning is that human language can express arbitrary levels of *complexity*. This is what makes it different from animal communication systems and for this, recursion that can embed constituents within constituents (either as tail-recursion or as CER) is responsible.

This thesis focuses on recursion that distinguishes FSGs and PSGs. Our training stimuli for both the human participants and the artificial neural network were centre-embedded artificial strings. We are aware that these do not model recursion of natural sentences entirely, since the between-phrase dependencies are missing. However, learning of the centre-embedded structure of strings and processing of the long-range dependencies present in these strings are prerequisites for parsing real CER.

2.2.4 Summary of the various meanings of recursion

The many meanings of recursion lead to various misunderstandings in the literature. One of them is claiming that recursion is responsible for the discrete infinity of language when the author obviously means recursion in production rules or recursion as in self-embedded structures. While it is true, that recursion in the global sense, as a general property of language as a computational system does make it possible to map sounds and meanings unboundedly, but it has little to do with the recursive nature of certain rewriting rules. The simplest proof of this point is that nonrecursive FSG rules can generate an infinite number of sentences too, if sentence length is unlimited.

The second misunderstanding stems from assuming that recursive rewrite rules generate recursive strings and these in turn must be parsed recursively. This is why the first generation of AGL experiments were methodologically flawed. In fact, recursive rules generate structureless strings, which can be parsed in numerous ways. The second generation of AGL experiments (including ours) partly solved the first half of this

[7] Simon Kirby, pers. comm.

problem (recursive rules → structureless strings) by establishing word-pairs by means of statistical co-occurrence, phonological cues or meaning. I say partly, because word-pairs represent within-phrase dependencies while between-phrase dependencies are still out of the picture. Nevertheless, artificial sentences in these experiments do have a centre-embedded structure, which is recursion in the “X within X” sense.

The recursive nature of parsing is a similar problem. In these experiments nothing makes sure that parsing is recursive in the sense as the production rules were recursive. What successful parsing proves is that the parser learnt that sentences have a centre-embedded structure. In some of our experiments, after training human participants were asked to formalize the learnt rule. Most of the successful participants provided a formula, something similar to $A_1A_2A_3B_3B_2B_1$ and expressed that this could be extended/generalized to longer or shorter sentences. This involves centre-embedding and infinity, which I think proves the recursive nature of the learnt rule. However, it still does not prove that online parsing was recursive in nature.

In fact, informal discussions with participants revealed that at least some of them parsed the sentences sequentially, continuously repeating the first half of the sentence until a pair turned up and then deleting the last word from the sequence held in memory. However, it might be the same with natural sentences too: A person understanding a sentence like “The man the dog bit yelped” could process it sequentially holding the noun phrase “The man” while processing “the dog bit” then returning to complete the outer phrase “The man yelped.” (Corballis, pers. comm.). This means that after reading the sentences in our experiment, even if one deduces a recursive rule, it is possible that the same person parses the sentences in a non-recursive way. In practice, it could be simpler to process them sequentially as it is true for natural sentences in natural language.

2.3 Recursion-only hypothesis

Recursion is a central issue of recent linguistics since the influential paper of Hauser, Chomsky, & Fitch (2002; henceforth HCF) where the authors introduced the distinction between faculty of language in the broad sense (FLB) and faculty of language in the narrow sense (FLN). “FLN includes the core grammatical computations” (abstract linguistic computations or narrow syntax) that are limited to recursion according to their *recursion-only hypothesis* (the name was coined later by Steven Pinker & Jackendoff, 2005). It interacts and interfaces with other systems of FLB. FLB includes FLN combined with the sensory-motor (phonetics/phonology), conceptual-intentional (semantics/pragmatics) and possibly other systems which make it possible for humans but no other animals to learn language without explicit instructions. It excludes other systems that are necessary but not sufficient for language (e.g., memory, respiration, digestion). The distinction between FLN and FLB is based on the uniqueness of their components: components of FLB are shared with other animals, while FLN is uniquely human and unique to language. They argued that FLN may have evolved for reasons other than language, i.e., it is an evolutionary spandrel and was not shaped by natural selection.

This paper not only generated a theoretical debate (W. T. Fitch, Hauser, & Chomsky, 2005; Jackendoff & Pinker, 2005; Steven Pinker & Jackendoff, 2005) but also stimulated empirical research on recursion using AGL tasks. Pinker and Jackendoff (2005) agreed in distinguishing between FLN and FLB, in dissecting FLB into sensorimotor, conceptual-intentional and grammatical components. They mainly criticised the recursion-only hypothesis and the assumption that language is not an adaptation. In their response Fitch, Hauser, & Chomsky (2005) pointed out that it was not language as a whole but only FLN that was supposed to be an evolutionary spandrel, and because of this, part of Pinker and Jackendoff (2005)'s critique is irrelevant.

More relevantly for the present thesis is the spur of empirical research that was stimulated by a paper from Fitch and Hauser (2004, henceforth FH); a paper that was allegedly misunderstood many times. This paper described an experiment which compared the performance of humans and tamarin monkeys in an AGL task. Strings were generated by an FSG and a CFG grammar, and the authors claimed that tamarin monkeys could learn FSG only, while human participants learned both. CFG strings conformed to counting recursion instead of real CER, due to the absence of word-pairs in the language. Moreover, word classes were differentiated by a very salient feature: Class *A* of words was spoken by a female and Class *B* of words was spoken by a male one octave lower in pitch.

Most misunderstandings stemmed from the fact that this paper was perceived as a direct continuation of the HCF paper. Not only because two of the authors are common but also because the HCF paper cited the F&H research as an unpublished manuscript. Although HCF never defined recursion in their paper, in most of the paper they asserted that this is the component of language that provides the means for discreet infinity that makes language open-ended with limitless expressive power. However, at the end of the paper (p. 1577) they used recursion as the procedure that distinguishes FSG and PSG:

> “[FSGs] are inadequate to capture any human language. Natural languages go beyond purely local structure by including a capacity for recursive embedding of phrases within phrases, which can lead to statistical regularities that are separated by an arbitrary number of words or phrases. Such long-distance, hierarchical relationships are found in all natural languages for which, at a minimum, a ‘phrase-structure grammar’ is necessary.”

Shortly after this quote, HCF cited F&H as an example for the comparative research that is needed for “exploring key differences between humans and animals relevant to FLN” (p. 1578). The focus of the F&H paper was recursion as a dividing line between FSG and PSG. Moreover, since they concluded that humans could learn CFG but monkeys could not, the logical inference was that they intended this experiment as a proof for the recursion-only hypothesis. Although F&H claimed several times that they did not even mention the word recursion in their paper (W. Tecumseh Fitch, 2010) they certainly implied it. “In addition to concatenating items like an FSG, a PSG can embed strings within other strings, thus creating complex hierarchical structures (‘phrase structures’), and long-distance dependencies” (p. 378). This is exactly the intended structure of strings on their first figure (reproduced here as Figure 5c). However, as Corballis pointed out, nothing established the dependencies between words thus sentences could be parsed by simply counting *A*s and *B*s. Moreover, Class *A* and *B* of words were differentiated by very salient features (pitch and sex of speaker).

There are six different scenarios for AGL experiments according to the type of word-pairs and word classes, see Table 3. If word-pairs are one-to-one, word classes only mean an order between *A* and *B* words. In other words ABs are ordered pairs, but word classes do not have any other meaning. If word-pairs are all-to-all (F&H design) it is not sensible to speak about pairs, just classes. In all scenarios, except for that of F&H statistical learning would be necessary, for which more habituation is needed.

Table 3. Possible experimental stimuli according to the mapping between the two classes of words and the word assignment to classes.

Possible scenarios		word assignment to classes	
		trivial	learnt
word-pairs	one-to-one	A) P&R[8]	B)
	few-to-few	C)	D) realistic
	all-to-all	E) F&H	F)

2.4 Outline

The following four chapters describe a connectionist model and three psycholinguistic experiments. These works are presented in chronological order to show the changing of methods and hypotheses in recursion research. This is a quickly developing area where hot topics cool down very fast and thus questions addressed in chapter 4 and 5 seem to be outdated now. However, the last chapter represents a very recent development, i.e. that semantics contributes to the learning of recursion, which has not been picked up yet by other research groups.

Chapter 3 and 6 were published in peer-reviewed journals and I decided to reproduce them here with only minor modifications. Although this decision increases the risk of redundancy within the thesis, I believe that chapters are more coherent and self-contained this way. Chapter 5 was presented as a poster during a conference (Recursion: Structural Complexity in Language and Cognition, 26-28 May 2009, University of Massachusetts, Amherst, USA); the poster is included in the Appendix (Section 9.1).

[8] (Perruchet & Rey, 2005a)

3 Recursion and stack

This chapter is based on a published paper of Fedor, Ittzés, & Szathmáry (2010). The original title was: *Parsing recursive sentences with a connectionist model including a neural stack and synaptic gating*. An earlier version of the model presented here was part of a book chapter(Fedor, et al., 2009). The motivation behind this work was to find a way with which the push-down automaton (that is known to effectively parse FSG and CFG strings) can be implemented neurally. This is not to say that humans parse sentences with a similar device, but to prove that the push-down automaton *can be* implemented in a biologically plausible way.

3.1 Abstract

It is supposed that humans are genetically predisposed to be able to recognize sequences of context free grammars with centre-embedded recursion while other primates are restricted to the recognition of finite state grammars with tail-recursion. Our aim was to construct a minimalist neural network that is able to parse artificial sentences of both grammars in an efficient way without using the biologically unrealistic backpropagation algorithm. The core of this network is a neural stack-like memory where the push and pop operations are regulated by synaptic gating on the connections between the layers of the stack. The network correctly categorizes novel sentences of both grammars after training. We suggest that the introduction of the neural stack memory will turn out to be substantial for any biological 'hierarchical processor' and the minimalist design of the model suggests a quest for similar, realistic neural architectures.

3.2 Introduction

Natural language is a fascinating phenomenon, very much in the focus of various disciplines, from linguistics proper to evolutionary biology (Derek Bickerton, 1990; Hauser, et al., 2002; Hurford, 2007; Maynard Smith & Szathmáry, 1995; S. Pinker, 1994). Although there is no general agreement on how to best characterize language, let alone its biological foundations, we follow the view that it is based on the critical combination of symbolic reference with complex syntax (Szathmáry, 2007). A crucial element of syntax is recursion (Corballis, 2007b; Hauser, et al., 2002). Two main types of recursion occurring in natural language are tail- or end-recursion (including left and right-branching recursion) and centre-embedded recursion (CER). An example of left-branching tail-recursion is (after the popular British nursery rhyme *The house that Jack built*):

(9) The rat squeaked.

(10)The cat killed the rat that squeaked.

(11)The dog worried the cat that killed the rat that squeaked.

If we represented word-pairs only, where word-pairs consist of two words with dependency between them (e.g., in sentence 3, these are: *dog-worried, cat-killed and rat-squeaked*), such sentences composed of three word-pairs can be described by the following rule:

(12)A_3 B_3 A_2 B_2 A_1 B_1,

where As represent nouns, Bs represent verbs, and words with the same index form word-pairs. Artificial sentences like this are characterized by the concatenation of coherent noun-verb pairs, and can be produced or parsed by simple iteration (M. H. Christiansen & N. Chater, 1999). Iteration is present in animal calls, like that of primates (e.g., Robinson, 1984; Zuberbühler, 2002) and songbirds (Eens, 1997). It was also proved that some animal species are able to infer the iterative rule from samples of artificial strings and generalize over novel strings (Robinson, 1984). Whereas it is possible to parse artificial sentences with tail-recursion by iteration, it is not true for natural sentences with tail-recursion, such as sentence (11), because in natural sentences apart from within-phrase dependencies (that establish word-pairs) between-phrase dependencies are also present. However, these are not represented by AGL tasks.

The előbb sentences [(10) and (11)] can be transformed to have centre-embedded structure:

(13)The rat that the cat killed squeaked.

(14)The rat that the cat that the dog worried killed squeaked.

Here, the general rule for three word-pairs is:

(15)A_1 A_2 A_3 B_3 B_2 B_1.

CER is claimed to be a general human capacity whereas it cannot be found in animal communication systems (W. Tecumseh Fitch & Hauser, 2004). It has not been proved either that any animal species is capable of learning CER in the laboratory. CER can be parsed by context-free grammar (CFG), which has higher generative power than finite-state grammar (FSG) which basically concatenates items (applies tail-recursion, Corballis, 2007b).

In natural languages, every noun has several possible verb pairs, and vice versa, every verb has several possible noun pairs. In the above examples semantic relationship connects the words: usually only rats squeak, not dogs or cats. However, if the other two word-pairs are swapped, the sentence still makes sense:

(16)The dog killed the cat that worried the rat that squeaked.

In a slightly modified version of the sentence, it is not possible to swap the word-pairs, because singular and plural words must be matched:

(17)Dogs worry the cat that kills rats that squeak.

As a result of this grammatical constraint (together with the semantic constraint), even if the words were mixed without grammatical structure, it would be easy to see the coherent noun-verb pairs. In artificial languages without semantics, if these dependencies between words are not established somehow, sentences could be represented by a simpler structure, A^nB^n, e.g. for n=3:

(18)A A A B B B.

This grammar is called counting recursion (M. H. Christiansen & N. Chater, 1999) because parsing of this kind of sentences is possible by counting As and Bs. If the number of As and Bs is equal and there is only one transition from As to Bs the sentence is correct (Corballis, 2007a, 2007b).

The structures that word-pairs imply are shown on Figure 5. In the case of tail-recursion, members of word-pairs are next to each other connected by local dependencies. Additionally, in sentences with CER there are word-pairs whose members are separated by other word-pairs thus they have long-distance (or long-range) dependencies. This implies a hierarchical structure compared to the linear structure of sentences with tail-recursion. The more levels this hierarchical structure has, the more words have to be remembered to be able to parse these sentences. In a six-word-long sentence with CER, the maximum number of words that has to be stored in memory is three and the first word has to be remembered until the presentation of the last one. In sentences with tail-recursion, members of word-pairs are presented shortly after each other; hence there is only one word that has to be remembered at a time. In counting recursion no individual word has to be stored in memory for grammaticality judgement, just the category of words passed and their quantity has to be remembered until the end of the sentence.

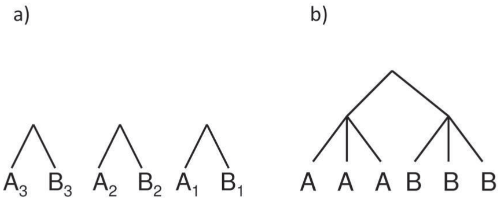


Figure 5. Tail-recursion (a), counting recursion (b) and centre-embedded recursion (c). ***A*** **and** ***B*** **represent word categories and As and Bs with the same index form word-pairs. Word-pairs imply only local dependencies in a) but also long-range dependencies in c). There are no word-pairs in b) (Corballis, 2007b).**

Some confusion resulted from using sentences of counting recursion in artificial language learning experiments and not differentiating them clearly from the more complex centre-embedded sentences. Fitch and Hauser (2004) claimed that their human participants were able to learn both FSG and CFG in a small artificial language, whereas cotton-top tamarins could learn only FSG. Since they did not establish dependencies between words their CFG sentences could be parsed by counting recursion. Likewise, Gentner, Fenn, Margoliash & Nusbaum (2006) claimed that their participants (starlings) learnt CER using the same kind of structures as Fitch and Hauser (2004). This in turn elicited some critique: Corballis (Corballis, 2007a, 2007b) called attention to the fact that the sentences used could be parsed by simple counting, while others (Perruchet & Rey, 2005b; de Vries, Monaghan, Knecht, & Zwitserlood, 2008) showed that in experimental situation similar to that of Fitch & Hauser (2004) even human participants used alternative strategies to solve the tasks.

Despite the controversy in artificial language experiments, there is a general agreement that centre-embedded structures are present in natural human languages, but not in natural animal communication systems. The question is, why. A simple answer would be that animals do not have Universal Grammar, but this leaves completely in the dark what the relevant biological differences could be. In fact there are serious doubts on the idea that abstract rules of Universal Grammar could ever get assimilated in the genome (Deacon, 2003; Wiles, Watson, Tonkes, & Deacon, 2005) but, in contrast, it cannot be doubted that some language-related genetic differences between humans and animals do exist. It is perhaps much more rewarding to enquire about the possible neuronal operations (procedures) that the brain could implement in order to handle language.

Connectionist models have increasingly been used to model empirical data across many areas of language processing (M. H. Christiansen & N. Chater, 2001). However, connectionist models aiming at parsing recursive structures are often restricted to counting recursion. Sun et al. (1998) implemented a hybrid system, in which a recurrent neural network was coupled to an external non-neural stack memory. After training with backpropagation, the system was able to infer a CFG from input. In another study, continuous-time recurrent networks without a stack can learn both context-free and context-sensitive languages in a prediction task, using backpropagation through time (Boden & Wiles, 2000). Since there were no long-range dependencies connecting words within the sentences, performance of these systems boiled down to counting (Rodriguez, Wiles, & Elman, 1999).

Other studies used input data conforming to real CER (as opposed to counting recursion) to train artificial neural networks. Elman (1991) trained a simple recurrent network (SRN) on multiclausal sentences which contained multiply-embedded relative clauses. The network achieved a high level of performance in predicting the next word in the sentences. In a related model Christiansen and Chater (1999) trained SRNs on recursive artificial languages. The behaviour of these networks was similar to human performance in that they reached higher performance in right-branching structures than in centre-embedded structures. In both studies backpropagation of error was used as a learning algorithm which is generally considered biologically implausible because it requires passage of information backward through synapses and along axons and because it uses error signals that must be precise and different for each neuron in the network (Mazzoni, Andersen, & Jordan, 1991; Randall C. O'Reilly, 1996).

Handling of hierarchical structures occurs at high speed during language production and comprehension, and it seems reasonable to assume that it requires specialized neural networks to do so (Fedor, et al., 2009). It is well known that parsing of CER can be solved very efficiently by a stack (push-down automaton), with the necessary pop and push operations (Hopcroft & Ullman, 1979). Thus it would be a step forward to present a neurally plausible simple stack architecture that could parse CER. Along this line, Chen and Honavar (Chen & Honavar, 1999) proposed an artificial neural network architecture for syntax analysis which is assembled from neural network components for lexical analysis, stack, parsing and parse tree construction. The stack in their model is a fairly complex system that is composed of five different modules that have specifically designed connections and the stack requires four sets of binary inputs. We aimed at constructing a more minimalist architecture for a neural stack that is more similar to a push-down automaton in its architecture.

In this chapter we will present a neural network which can be trained to parse sentences with tail-recursion and CER. Three key features of the model are: a) absence of backpropagation, b) a crucial role for synaptic gating, c) and a neurally implemented stack. These components of our model are not new, but the combination of these features is unprecedented – this is what makes this model very effective and minimalistic.

3.3 Methods

3.3.1 Grammars

We composed input sentences according to two types of recursion, namely tail-recursion and CER. Words were 0/1 binary strings, where there was only one 1 in each word (all the other digits are 0). Words were randomly divided into two classes, *A* and *B*. Each word from Class *A* had exactly one (randomly chosen) pair from Class *B*, and vice versa. Sentences (12) and (15) give examples for six-word-long sentences with tail-recursion and CER, respectively. Since no word occurs twice in a sentence, 8*7*6=336 sentences could be generated for each grammar.

Additionally, random ungrammatical sentences were also generated. These sentences were also composed of three *A* words and three *B* words and always started with an *A*, just as grammatical sentences, but did not conform to any of the above rules.

3.3.2 Architecture and functioning of the network after successful training

The neural network consists of the following main modules: input layer, stack, predictor, two push-pop neurons and a decision neuron (Figure 6). The input layer receives one word at a time from the sentence. In the case of a 16-word vocabulary, the input layer has 16 units (neurons), where each unit corresponds to a single word. Second, there is a clocked stack (for a clocking mechanism see Hjelmfelt, Weinberger, & Ross, 1991) with three layers, where every layer of the stack consists of 16 neurons. Adjacent neurons within a column of the stack are connected bidirectionally to each other, with a weight of 1. The third component, called the predictor, tries to predict the next word in the sentence based on the word that is stored in the top layer of the stack. The push-pop neurons have input connections from the predictor and the input layer. They basically compare the two, and if they store the same words (which means that the prediction was correct) signal +1, if they store different words (or if the predictor is empty) signal -1. The output connections of the push-pop neurons perform neural gating on the synapses of the stack: these connections modulate the synapses directly by enhancing or inhibiting them (i.e., the push-pop neurons are the so called ‘gatekeepers’). One of the push-pop neurons is an excitatory neuron which is connected to each upward synapse in the stack with positive weights (i.e., gating acts in a permissive fashion; Katz, 2003) and the other one is an inhibitory neuron, which is connected to each downward synapse in the stack with negative weights (absolute suppressive gating). The signal that arrives to a synapse from a gating neuron is the product of the activation of the gating neuron and weight on the synapse of the gating neuron, just like with any other neurons. The difference is that a negative signal from a gating neuron blocks the synapse it is connected to, while a positive signal from a gating neuron makes it possible for the synapse to work. As a result, if the prediction was correct and the push-pop neurons signal +1, downward connections will be inhibited and upward connections will be enhanced in the stack, hence upward connections will predominate, and each layer will take the value of the layer bellow it (a pop action). In this case, the bottom layer becomes empty. On the other hand, if the prediction was not correct and the push-pop neurons signal -1, upward connections will be inhibited and downward connections will be enhanced in the stack, such that the downward synapses will predominate and every layer will take the value of the layer above it (a push action). In this case, the top layer takes its value from the input. Lastly, there is a decision neuron, which is connected to the top layer of the stack and signals only if there is a word stored on the top layer of the stack. The signalling of this neuron can be considered as the decision of the network on the grammaticality of the sentence: signalling means that the sentence encountered so far was ungrammatical, while 0 output means that the sentence is grammatical.

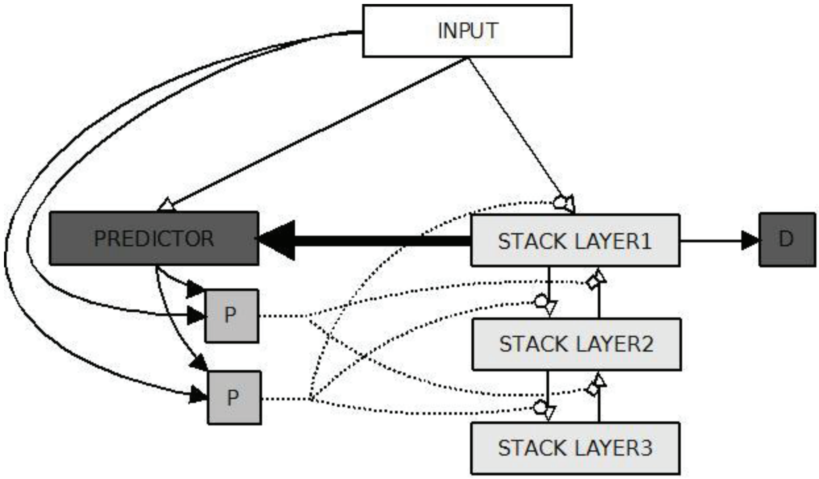


Figure 6. Architecture of the proposed neural network model. The input layer receives words from sentences one-by-one. The stack is represented with three layers with bidirectional inter-layer connections. The predictor layer tries to predict the next word based on the word that is stored at the top of the stack (stack layer 1). There are two push-pop neurons (P) that has gating connections (dashed lines) on the inter-layer connections of the stack. Inhibitory gating connections are marked by a circle at the end and excitatory gating connections are marked by a diamond at the end. The decision neuron (D) gives grammaticality judgement on the sentence. Copying connections that are not trained are indicated by empty arrows. Synapses between the top of the stack and the predictor are indicated by a thick arrow and trained by the Hebbian learning rule. All other synapses are trained by the perceptron learning rule.

Now, let us see how the whole network with a three-layer-deep stack is supposed to parse a six-word-long sentence after learning. First, the predictor tries to predict the first word from the top of the stack, but since the stack is empty at the beginning, the predictor will have no prediction. Next, the first word arrives to the input layer and then the push-pop neurons compare the input with the prediction. Since prediction is unsuccessful (there is no word on the predictor), the push-pop neurons perform a push action on the stack and the top layer of the stack becomes occupied by the first word. This triggers the decision neuron, which will signal that the string encountered so far is ungrammatical. In the case of a sentence with tail-recursion, the next word depends on the previous one. At the beginning of the next cycle, the predictor predicts the next word based on the previous word which is stored on the top of the stack. If the prediction is correct, the push-pop neurons will perform a pop action on the stack, which will become empty again. This will suppress the signalling of the decision neuron which means that the sentence encountered so far is grammatical. The same push-pop actions are repeated with the next two word-pairs until the end of the sentence. To measure the performance of the network we can detect its predictions for the following words or its decisions on the grammaticality of the sentence. During the processing of a sentence with tail-recursion, the network is able to predict every second word (other words cannot be predicted, hence maximum performance is 50%) and decides that the sentence is incorrect three times: after the first, third, and fifth word. Note, that independently of the length of the sentence, substrings of a grammatical sentence with even number of words are in fact grammatical in the case of tail-recursion.

In the case of an ungrammatical sentence, one or more words do not have a pair, which means that more than three words are unpredictable in the sentence. This results in more than three push actions, which means that the stack is not empty at the end of the sentence and the decision neuron will signal that the sentence was ungrammatical.

Parsing grammatical sentences with CER also involves equal number of push and pop actions; hence the stack is empty at the end of a grammatically correct sentence. The difference is that there are three push actions until the predictor can finally predict a word. Prediction is always based on the word that is stored on the top of the stack, and after three unsuccessful predictions on the first three words, these words are stored in the three layers of the stack, with the third word being on the top. The fourth word is predictable from the third word, which means that the push-pop neurons will perform a pop action on the stack, after which the second word will be on the top. The fifth word can be predicted from the second word, which means another pop action, and finally, the sixth word is predicted from the first one. As in the case of tail-recursion, the stack is empty after the presentation of a grammatically correct sentence.

It can be seen that in the case of a sentence with tail-recursion, only the top layer of the stack is used, whereas for parsing a six-word-long sentence with CER, three layers are used. Generally, the number of stack layers required for parsing a centre-embedded structure is half of the number of words in the sentence. If there are fewer layers, the network will categorize sentences with CER as ungrammatical. Note that tail-recursion can be parsed without push-pop neurons and stack if you use simple copying from the input layer to a one-layer memory instead of a push action and deletion of the memory instead of a pop action. The difference between animals that cannot parse CER and humans can be that the former lack the stack and the gating mechanism, without which only local dependencies between words can be parsed.

It can be argued that the stack architecture in this form cannot explain why deeper embeddings are harder to process for humans. With this solution two levels of embedding (6-word-long sentences) are processed perfectly, whereas three or more levels are impossible. However, if we realize that every neural computation is prone to errors, we will see that the stack architecture with many layers also shows graceful degradation in performance as the level of embedding increases. If every push or pop operation in the stack has a small probability to result in imperfect transmission of information from one layer to another, then the more embedding the sentence have the more probable is its faulty parsing.

3.3.3 Training

While the architecture of the network described above is hand-crafted, its synaptic weights develop during training. For the training we randomly chose a subset (the learning set) from the grammatical sentences of either type of recursion. Training consisted of presenting the learning set several times and modifying the weights of the network. Testing was performed on the rest of the grammatical sentences that the network has not encountered before randomly mixed with ungrammatical sentences. During testing no weight change occurred. The performance of the network was measured by its predictions for the following words in grammatical sentences during training and testing and its decisions on the grammaticality of the sentences at the end of the sentences. Note, that theoretically the maximum performance for prediction is 50% in grammatical sentences (obviously it was not measured for ungrammatical sentences). Grammaticality judgement during testing measures if the network can differentiate grammatical from ungrammatical sentences, while during training it is not very informative, since there were only grammatical sentences in that phase.

Different learning rules were used to modify the weights of the network. For the weights between the top layer of the stack and the predictor layer, a simple Hebbian learning rule was used. After the predictor layer tried to predict the next word and the push-pop neurons compared the prediction with the next word on the input, the input was copied to the predictor. Then learning occurred in this time step by increasing the synaptic weights between those neurons that were activated:

if $N_t = 1$ *and* $N_p = 1$ *then* $W_{tp} = W_{tp} + r$,

where N_t is a neuron on the top of the stack, N_p is a neuron on the predictor layer, W_{tp} is the synaptic weight between them and r is the learning rate (r was set to 0.01).

For modifying the synaptic weights of the push-pop neurons coming from the predictor and the input layer and the synaptic weights of the decision neuron coming from the top of the stack, the perceptron learning rule was used with threshold transfer function (Dayan & Abbott, 2005). This learning rule modifies the weights and the threshold to reach an output that is closer to a precalculated desired output. For this only local information is used: the activation of the input and the output layer (e.g., in the case of the decision neuron the input is the top of the stack) and the synaptic weights:

$$W = W + h*(O - O') * I \quad \text{and} \quad T = T - h*(O - O'),$$

where W is the weight matrix between the input and the output layer, I is the input, O is the desired output, O' is the actual output, T is the threshold for the transfer function and h is the learning rate (h was set to 0.1). The desired output for the push-pop neurons is 1 (pop) if the prediction was correct (i.e., if the input layer and the predictor layer has the same activation pattern) and -1 (push) if the prediction was incorrect. For the decision neuron, the desired output is 0 if the top of the stack is empty and 1 otherwise. (Note that starting with nonzero random weights, the decision neuron automatically works well without learning.)

Weight modification occurred online, i.e. after the presentation of each word in the case of both learning rules. (We tried batch learning too, where weight modification occurs after the presentation of the whole training set. The network basically reached the same performance, however, we find it less realistic, and hence we used online learning for generating the figures in this paper.)

Those weights that copy activation from one layer to another were not trained, but set to the desired values from the beginning: between the input and the predictor, between the input and the top of the stack, and the weights between the layers of the stack. It might seem to be quite artificial that the weights of the stack are not trained but are precalculated. However, since it is a very simple structure it would be easy to train in an extended version of this model, just as the other copy weights.

3.4 Results and conclusions

3.4.1 Learning performance

Figure 7 a) and b) show the performance of the model during several training sessions and a test session averaged over 10 runs in the case of tail-recursion and CER, respectively. Performance is measured by the correctness of grammaticality judgement at the end of sentences (Decision) and by the correctness of the prediction for words during sentences (Prediction). Black data points represent performance during training while the last white data points represent performance during testing. Training was performed on a randomly chosen subset of the 336 grammatical sentences, while testing was performed on the rest of the grammatical sentences mixed with ungrammatical sentences. Note, that the theoretical maximum for prediction performance is 50% in the case of grammatical sentences for both grammars (it was not measured for ungrammatical sentences).

In the case of tail-recursion a training set composed of 10 grammatical sentences was usually enough for the network to generalize and reach perfect or almost perfect performance on novel sentences. For this about 5 training sessions were needed. In the case of CER, 30 training sentences presented for 11-12 training sessions were needed to reach the same performance. For both grammars, perfect performance on novel sentences is possible provided that every word-pair is presented during the training sessions. There is no generalization on the level of word-pairs; it is simply not possible since words are paired randomly. However, the network successfully generalizes on the level of sentences as can be seen from its performance on novel test sentences.

The network can also be trained with a mixed set of sentences conforming to tail-recursion and CER. With 30 sentences, only about 6-7 training sessions are needed to reach perfect performance which indicates faster learning than with CER sentences only. This is quite intuitive: tail-recursion seems easier to learn than CER since words forming a word-pair are presented immediately after each other. For the successful parsing of the grammars both memorizing the word-pairs and recognizing the particular structure is necessary. Since finding the words that depend on each other seems to be easier in the case of tail-recursion, we predict that it would help humans to learn CER if sentences were mixed with tail-recursive sentences.

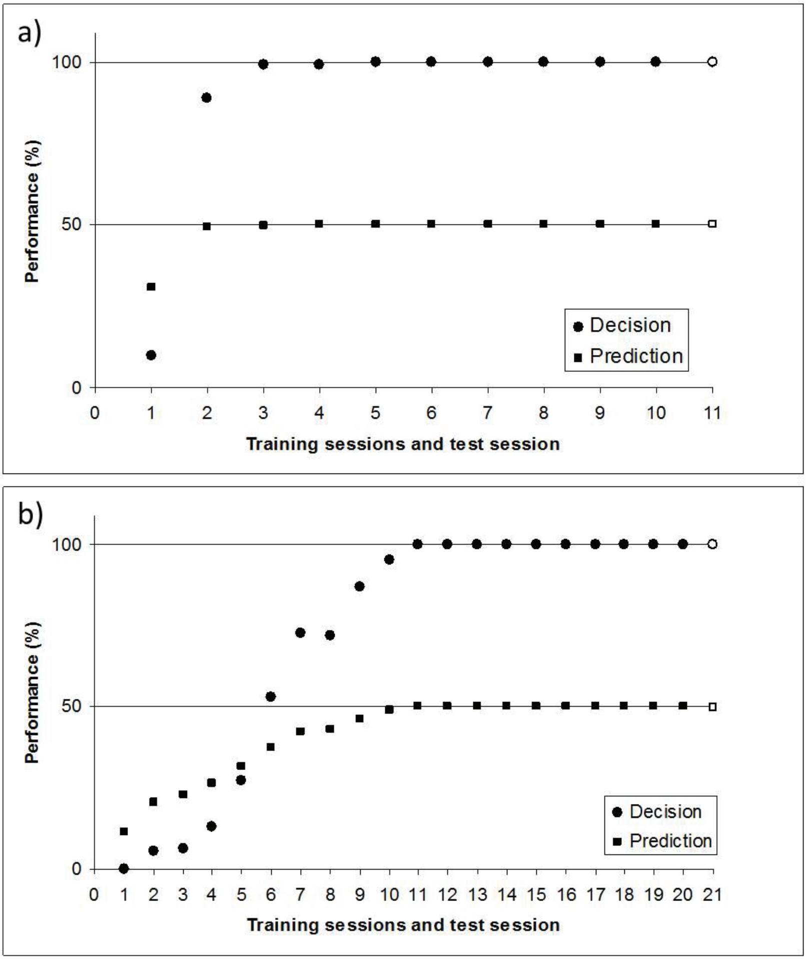


Figure 7. Performance of the neural network model on a) tail-recursion and b) centre-embedded recursion averaged over 10 runs. Performance is measured by the correctness of grammaticality judgement at the end of sentences (Decision) and by the correctness of the prediction for words during sentences (Prediction). Black data points represent performance during training while the last white data points represent performance during testing. Training was performed on 10 and 30 randomly chosen grammatical sentences in the case of tail-recursion and centre-embedded recursion, respectively. Testing was performed on the rest of the grammatical sentences mixed with ungrammatical sentences.

3.4.2 Extensions

Since the performance of the network is based on the successful learning of word-pairs coupled with the push-pop operations of the stack, it can learn any language that is based on the balanced pairing of words in sentences. One example is the Dyck language of balanced parenthesis, which can be thought of as a mix of tail-recursion and centre-embedded recursion. E.g.: strings like *aaabbaabbabb* can be parsed by the model after learning that *a* and *b* are pairs. Another example is the palindrome (mirror) language; sentences like *abccba* can also be parsed by the network. The difference with centre-embedded recursion is that in this case word-pairs consist of two identical words.

The network cannot learn counting recursion in this form, since it has no module that would learn to categorize words to Class *A* and Class *B*. However, if we inserted a module that was able to categorize words on the predictor and on the input layer, it would make it possible to parse sentences with counting recursion too.

3.4.3 Conclusions

The main features of the neural network implemented here is the neurally implemented stack operated by gating neurons[9]. While it is well known that recursive sentences can be parsed by a symbolic stack, to our knowledge, there was no simple neural implementation of this structure until now. Symbolic models that could not be neurally implemented can be ruled out for being implausible (Morten H. Christiansen & Nick Chater, 2001). What we show here is that the stack indeed can be neurally implemented, and it is quite simple provided that the push-pop operations are guided by gating connections.

We believe that gating will be found crucial for hierarchical tasks, just as for complex cognition in general (Gisiger, Kerszberg, & Changeux, 2005; R. C. O'Reilly, 2006). The fact that it has readily evolved in a reinforcement-learning task in a simulated honeybee neural network (Soltoggio, Dürr, Mattiussi, & Floreano, 2007) supports this idea. We suggest that the introduction of the neural stack memory (push-down automaton) will also turn out to be substantial for any biological 'hierarchical processor'. This is not to say that it is just a neural stack that is crucial for language, neither do we suggest that the stack architecture proposed here exists in a clean, isolated form in the

[9] The stack follows a design borrowed from the chemical literature (Hjelmfelt et al., 1992) that rests on gating.

brain, but it is likely that similar networks are embedded in the wider, language-related network context.

The performance of our network naturally depends on the depth of the stack, and as such it can be replaced by a finite-state automaton (Hopcroft & Ullman, 1979). However, in this sense human parsing ability is also limited: no person can parse sentences with arbitrarily many levels of embeddings (S. Pinker, 1994). The likely hierarchical processor (maybe even supramodal) in humans with normal development is Broca's area (Bahlmann, Heim, Schubotz, & Friederici, 2006; Tettamanti & Weniger, 2006). Sadly, we know next to nothing about the relevant 'internal wiring' of this area: we propose that it is likely to contain a neural stack, wherein gating will be found important.

It would be premature to contemplate about the origin of stack-like neuronal systems in evolution and development. However, there seem to be two possible scenarios: either stacks are hard-wired (genetically coded) in our brain, or we are born without them and the plasticity of our brain (under genetic control) makes us 'ready' to organize stacks during development. We think that the second scenario is more plausible but future work is needed to resolve these issues.

4 Recursion and silent gaps

4.1 Introduction

Perruchet and Rey (2005b; from now on P&R) replicated the experiment of Fitch and Hauser (2004; from now on F&H) to test what participants base their decisions on when discriminating grammatical from ungrammatical strings: is it the true understanding of embedded hierarchical structures, or is it a more superficial knowledge about the sound patterns of strings? Suspicion emerged that participants did not learn CER in the F&H experiment, out of several reasons:

1. There were no word-pairs in the artificial sentences, thus learning CER was not necessary to successfully parse strings; the simpler counting recursion was sufficient.
2. Word classes were indicated by very transparent cues: Class *A* words were uttered by a female voice and Class *B* words were uttered by a male voice one octave lower in pitch
3. The habituation phase lasted only 3 minutes for humans.

The main modification in P&R's experiment was that parallel to the word classes that were indicated by different sounds (high and low pitched) they used word-pairs according to the advice of Corballis (2007a). In this way, the grammar they used could be described by the rule $A_1A_2A_3B_3B_2B_1$ as opposed to AAABBB as in F&H. During familiarization, grammatically correct sentences conformed to both the acoustic pattern (HHHLLL, where H is high pitched and L is low pitched as in F&H; in P&R there were no sex differences, because all voices were generated by software) and the word-pair pattern (123321). For testing, they composed ungrammatical sentences that violated either the acoustic pattern or the word-pair pattern or both (see Table 4).

Table 4. Structure of test sentences in P&R

Grammatical Structure (Center-Embedding)	*n*	Acoustic Pattern (Pitch Variation): Violation	Acoustic Pattern (Pitch Variation): Consistent
Violation	2	**A1** A2 **B1** B2	**A1** **A2** B1 B2
	3	**A1** A2 **A3** B2 **B1** B3	**A1** **A2** **A3** B2 B1 B3
Consistent	2	**A1** A2 **B2** B1	**A1** **A2** B2 B1
	3	**A1** A2 **A3** B3 **B2** B1	**A1** **A2** **A3** B3 B2 B1

Note—Bold and underlined characters = high pitch; normal characters = low pitch.

Since the aim of the experiment was to test whether participants based their decision about the grammaticality of sentences on the salient class distinction feature or on the word-pairs, P&R tried to keep the same all but one important parameter: the existence of word-pairs. However, as can be seen in Table 5, they have also changed some other, supposedly unimportant parameters.

Their participants were French students, whereas in the F&H experiment they were supposedly English speakers (although it is not specified in the paper). While we suppose that different nationalities possess basically the same abilities to parse sentences, we cannot be sure that the same artificial syllables are similarly familiar to participants of different mother tongues. First of all, one has to make sure that the combinations of sounds used in the syllables are allowed by the phonotactic rules of a language. This criterion was met by both studies. Second of all, one has to make sure that all syllables used are easily distinguishable for the participants. For this reason P&R changed some of the syllables: they replaced *yo* by *ro*, *wu* by *vu*, and and *pa* by *sa*. Third, the familiarity of the same artificial syllables can be different for participants of different mother tongues. We investigated this issue in another experiment described in chapter 6.

Generation of the speech sounds was also different. While F&H used recorded human voice to compose the sentences, P&R used a speech synthesizer. While speech synthesizer voice can be very strange P&R must have supposed that it would not influence the discrimination of syllables. F&H recorded *A* and *B* Groups of words by a female and male speaker, respectively. They also transformed the sounds so that the female voice was an octave higher in average. P&R used only pitch differences to help discriminating Class *A* and *B*. According to them, the difference was salient enough in this way too. And last, while in F&H there was a silent gap between most words in a sentence, in P&R's experiment sentences were continuous. Although it was not specified in the paper, but based on the sample stimuli provided online, it seems that word length and gap length were not controlled in F&H's experiment. I analysed the stimuli and found that word length was in the range of 390-660 ms (average = 510 ms), and gap length was 0-560 ms (average = 70 ms). In P&R word length was similar (450 ms), but there was no silent gap between words.

While it can be supposed that most of the above changes did not have a major effect on AGL, there is a reason to suspect that the last one did. It is well known that continuous speech poses higher processing demands than speech with gaps between words. In an experiment of Pena, Bonatti, Nespor, & Mehler (2002) adults were unable to discover regularities in an artificial language when the streams of speech were presented continuously, whereas they could learn the same patterned relations when subliminal gaps were inserted between words. In another experiment (Newport & Aslin, 2004) when stimuli were presented continuously adults were able to learn dependencies (associations) among adjacent syllables, but not among non-adjacent syllables. Since centre-embedded sentences are characterized by long-range non-adjacent dependencies (only the middle word-pair is adjacent), it is plausible to assume that learning CER is influenced by the absence of between-word gaps.

To test this hypothesis, we designed an experiment in which we replicated P&R's habituation and testing methods, with the main difference being that in our sentences there were 250 ms silent gaps between words. Word-pairs were present as in P&R. Other auditory characteristics of sentences made the stimuli more similar to that of F&H: words were recorded and played back instead of being synthesized by software; and were differentiated by not only pitch differences but also by the sex of the speaker.

Table 5 Comparing the methods of F&H, P&R and the present study. (F&H tested human participants and tamarin monkeys on tail-recursive and centre-embedded sentences. Here, I refer to the experiments with centre-embedded sentences on humans only.)

	F&H	P&R	Present study
Participants	10 undergraduate students, all female; supposedly English speakers	32 undergraduate students, all native French speakers.	25 participants, mainly undergraduate students, all but one native Hungarian speakers
Vocabulary	Class *A*: ba, di, yo, tu, la, mi, no, wu. Class *B*: pa, li, mo, nu, ka, bi, do, gu.	Class *A*: ba, di, *ro*, tu, la, mi, no, vu. Class *B*: *sa*, li, mo, nu, ka, bi, do, gu.	ba, di, jo, tu, la, mi, no, vu, pa, li, mo, nu, ka, bi, do, gu. Class *A* and *B*, and word-pairs were randomly assigned for each participant.
Speech voice	Speech syllables were recorded and then played back.	Speech was synthesized using the MBROLA speech synthesizer with the FR4 diphone database.	As in F&H.
Class distinction	Different syllables were spoken by a female (*A*) and a male (*B*) and were differentiated by voice pitch (1 octave difference), phonetic identity, average formant frequencies, and various other aspects of the voice source.	*A* syllables were set at a fundamental frequency of 240 Hz, and *B* syllables at 80 Hz.	As in F&H.
Word-pairs	No word-pairs.	The pairing between *A* and *B* syllables was	The pairing and grouping of syllables

		arbitrary and differed for each participant.	was arbitrary and differed for each participant.
Temporal characteristics of sentences	Based on the wav file of 8 sentences in the supplementary material, silent gaps between words and word length was not controlled. Average word length: 510 ms, average gap length: 70 ms.	No silent gaps between syllables. Sentences were separated by 3400 ms silent period.	Word length: 500 ms. gap length 250 ms. Sentences were separated by 1500 ms silent period.
Instructions	"Students were asked simply to listen to the sounds for three minutes, and then were asked to rate a set of novel sounds, stating simply whether the pattern of each novel sound was the same as or different from the previous set, by pressing a button on the computer screen. No feedback was given."	"[Participants] were told that they would listen to sequences of sounds for 3 min and that they would be asked to answer questions about the sounds at the end of the presentation. They were asked to avoid engaging in analytic problem-solving processes." "After familiarization, the participants were told that they would be presented with a set of novel auditory strings and that they would have to judge, for each one, whether the pattern was the same as or different from the pattern of the	Instruction before habituation: "You will hear 3 minutes of recording of artificial sentences. Please listen carefully!" Instructions before testing: "Now, you will hear 16 more sentences. Try to decide one-be-one, whether their pattern is consistent with that of the sentences that you were listening to in the previous 3 minutes. Mark your answers with C (consistent) or D (different) letters!"

		strings heard previously. The experimenter noted the participants' verbal responses."	
Habituation	Less than 3 min of exposure; 30 sentences.	Approximately 3 min.; 32 sentences: 16 four-word-long and 16 six-word-long sentences.	As in P&R.
Test	8 tail-recursive and 8 centre-embedded sentences.	16 test sentences as shown in Table 4.	As in P&R.

4.2 Materials and Methods

4.2.1 Participants

There were 25 participants, mainly undergraduate students, 10 male and 15 female. Average age was 22.5 years (range: 19-32 years). All but one participant were native Hungarian speakers. There was one English-speaking participant, to whom the instructions were translated.

4.2.2 Stimuli

We used the same vocabulary as F&H with the only change being that *wu* was replaced by *vu* (*w* is uncommon in Hungarian). For the purpose of recording the words, *yo* was spelled *jo* in the script of the speakers to approximate Hungarian orthography but the spoken phonemes were basically the same. Thus the 16 words (consonant-vowel syllables) were: ba, di, jo, tu, la, mi, no, vu, pa, li, mo, nu, ka, bi, do, gu. While in F&H and P&R the first eight syllables were always assigned to Class *A* and the rest to Class *B*, we randomly assigned words to classes for each participant to provide better randomization. As in P&R words were randomly paired for each participant in a way that each word from Class *A* had exactly one pair form Class *B*.

Words in Class *A* were spoken by a female voice and words in Class *B* were spoken by a male voice. The voice was recorded, and then words were modified with Audacity software to have equal length (500 ms) and so that words spoken by the female were one octave higher in pitch. We composed sentences for habituation and testing in the following way. We generated so called sentence patterns from numbers from 1 to 16, than we substituted the numbers with recorded words according to their grouping, pairing and ordering individually for each participant in accordance with their unique vocabulary. These patterns and the order of these were the same for all participants. In this way, the grammatical structure of the stimuli was the same for all participants, while the surface form differed, preventing the possibility that judgments were affected by a priori perceptual biases as in P&R.

First, we generated the first half of the sentences randomly. Then, we composed the second half of the sentences according to the grammar for the habituation sentences and with some systematic errors for the test sentences. The habituation set consisted of 16 six-word-long and 16 four-word-long sentences, all consistent with both the acoustic pattern (FFFMMM and FFMM, where F=female voice and M=male voice) and the word-pair pattern of the centre-embedded rule ($A_1A_2A_3B_3B_2B_1$ and $A_1A_2B_2B_1$).

We composed our test sentences according to three factors (length, acoustic pattern and word-pair pattern or grammaticality), as in P&R (Table 4). There were two examples of each sentence type in the test set to total up to 16 sentences. The final order of our test sentences can be seen in Table 6. Sentences violating the acoustic pattern were characterized by alternating female and male voices (FMFMFM for six-word-long sentences and FMFM for four-word-long sentences). Word-pair pattern violations were generated according to patterns $A_1A_2B_1B_2$ and $A_1A_2A_3B_2B_1B_3$ in four-, and six-word-long sentences, respectively.

The most important difference between our methods and those of P&R is that there were 250 ms silent gaps between the words of a sentence. The silent period between sentences was 1.5 sec. Sentences were played on a laptop computer through earphones in a silent room.

Table 6. Order of test sentences for each participant. V = violation, C = consistent. Note that although the structure of sentences was the same for each participant, the surface form differed due to the different arrangement of vocabularies.

Number	Word-Pair Pattern	Acoustic Pattern	Length
1	V	V	6
2	C	V	4
3	C	C	6
4	V	C	6
5	C	V	6
6	C	C	4
7	V	C	4
8	V	C	6
9	V	V	4
10	V	V	6
11	V	V	4
12	V	C	4
13	C	V	4
14	C	V	6
15	C	C	4
16	C	C	6

4.2.3 Procedure

Before habituation participants were given the following instructions: "You will hear 3 minutes of recording of artificial sentences. Please listen carefully!" [10] Habituation consisted of listening to 16 six-word-long sentences (about 4.25 sec each) and 16 four-word-long sentences (about 2.75 sec each) which took about 2 minutes and 40 seconds.

Testing occurred immediately after habituation. Participants received a sheet of paper with instructions and numbers from 1 to 16 on it. Instructions were the following: "Now, you will hear 16 more sentences. Try to decide one-be-one, whether their pattern is consistent with that of the sentences that you were listening to in the previous 3

[10] In Hungarian: *Most egy 3 perces felvételt fogunk lejátszani neked, mely mesterséges mondatokból áll. Kérlek, hallgasd figyelmesen!*

minutes. Mark your answers with C (consistent) or D (different) letters!”[11] The experimenter asked the participants whether they understood the instructions and explained them further if it was necessary. Participants listened to the test sentences and recorded their own answers on the paper.

4.3 Results

Figure 8 shows the percentage of “Different” answers given by participants to the different types of test sentences (as a percentage of the number of sentences of a given type). An ANOVA with Tukey-Kramer multiple comparisons post test revealed that there are significant differences between variables represented on the figure (number of “Different” answers to sentences conforming to or violating acoustic pattern and word-pair pattern independently), $F = 41.881$, $p < .0001$. Participants gave about the same number of “Different” answers to word-pair pattern consistent and inconsistent sentences ($q = .8332$, $p > 0.05$), which means that the word-pair pattern probably did not influence their judgments. On the contrary, there were significantly more “Different” answers to acoustically incorrect sentences than to acoustically correct sentences ($q = 15.830$, $p < .001$). This means that participants could learn the acoustic pattern but not the word-pair pattern during habituation and based their decisions on the former during testing.

To make our results directly comparable to that of P&R, we performed the same kind of statistical analysis that they did. It was a repeated measures ANOVA with no between subject factors and no covariates. The within subject factors were acoustic pattern (with 2 levels: consistent and violation), word-pair pattern (with 2 levels: consistent and violation) and the length of sentences (with 2 levels: 4 and 6 words). The variables were the number of "Different" answers from each participant for the different type of test sentences. Each variable could take the values: 0, 1, and 2 in each participant.

There was a main effect of acoustic pattern, $F(1, 24) = 47.956$, $p = .000$, but word-pair pattern did not have an effect, $F(1, 24) = .896$, $p = .353$, just as in P&R. We also had a main effect for sentence length, $F(1, 24) = 5.651$, $p = .026$, while P&R did not. In their analysis there was an acoustic pattern × sentence length interaction. In our case, there were no significant interactions, however, the acoustic pattern × sentence length

[11] In Hungarian: *Most még 16 mondatot fogsz hallani. Döntsd el róluk egyenként, hogy mintázatuk megegyezik-e az előző 3 perces felvételen hallottakéval! Válaszodat jelöld K (különböző), vagy E (egyező) betűvel!*

interaction was close to being significant, $F(1, 24) = 3.332$, $p = .080$. The pattern of results is depicted on Figure 9.

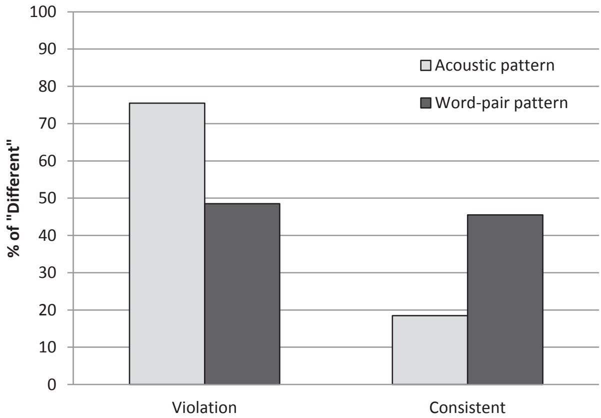


Figure 8. Percentage of "Different" responses given by participants, as a function of the well-formedness of the test sentences with regard to their acoustic pattern and word-pair pattern.

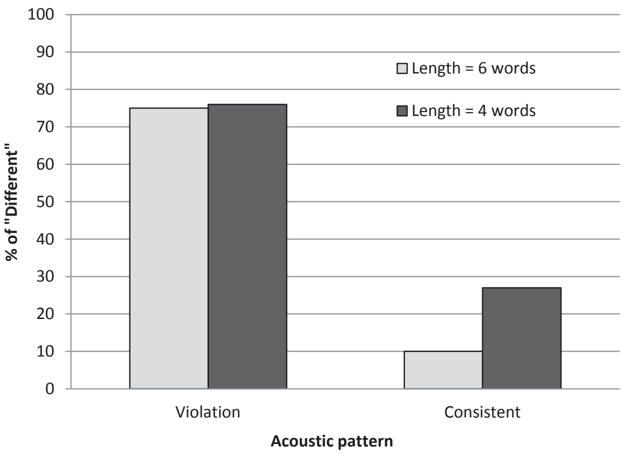


Figure 9. Percentage of "Different" responses given by participants, as a function of the well-formedness of the test strings with regard to their acoustic pattern, according to the number of words in the sentences.

4.4 Conclusions

We replicated the experiment of P&R, with the main difference being that we composed our stimuli with between-word gaps while their sentences were continuous streams of consonant-vowel syllables. In their experiment participants could not learn CER in three minutes of habituation. We hypothesized that the reason for this could have been that continuous speech demands higher processing capacities than sentences where the word boundaries are obvious, as in F&H.

Our hypothesis was not verified: our experiment gained very similar results than that of P&R, i.e., gaps between words did not make it possible for participants to learn CER during 3 minutes of habituation. This reinforces P&R's idea that human participants in F&H's experiment probably categorized sentences according to their acoustic patterns.

The acoustic pattern in these experiments (F&H, P&R and ours) conforms to the A^nB^n rule which is not the same that is present in natural languages, since it lacks not only the between phrase dependencies but also the within-phrase dependencies, i.e., word-pairs. Word-pairs in fact were not present in F&H's stimuli, but they were in the other two studies. Results show that even when word-pairs were established, they did not influence the judgments of participants. Of course, it does not mean that human participants cannot learn CER. It could mean that the acoustic pattern of sentences is just so much more prominent than the word-pair pattern that it distracts participants. Another possibility is that three minutes of habituation is just too short for humans to learn CER.

In the next chapter I describe an experiment where I tried to get around both of these problems by providing more habituation, less distracting word classes along with a supervised learning paradigm to enhance the probability that participants learn CER. Additionally, I also tested whether the difficulty lies in working memory demands.

5 Recursion and working memory

5.1 Introduction

Our previous experiment showed that human participants could not learn the centre-embedded rule in artificial sentences during 3 minutes of habituation. The main difficulty of the task is the following: participants have to realize that there are word-pairs, because without this knowledge, they cannot recognize the centre-embedded pattern. However, without recognizing the centre-embedded pattern (or at least parts of it), there is no way to know which words are pairs. If they first realize that there are word-pairs, they can memorize them, and then look for the rule with their help. Vice versa, if they somehow figure out the centre-embedded pattern first then they can use this knowledge to learn the word-pairs. Either way, they have to master both to successfully solve task. In this experiment we tried different stimuli and learning paradigm to see whether these changes make it possible to learn CER.

Human infants, while being naturally immersed in a fully complex language environment, are also exposed to Motherese, or child-directed speech, which is characterized by shorter sentences, simpler grammar and restricted vocabulary (Falk, 2004). During development, at around two-years of age, children go through a two-word stage when a large portion of utterances refer to “who does what”, i.e. in these simple sentences one of the words is the agent and the other one is the action (Edwards, 1973). By the time they try to understand centre-embedded sentences, they probably already know most of the words and they know usually who does what, which can help them figure out the structure of these complicated sentences.

Of course, the aim of AGL experiments with adults is not to directly model how infants learn language, since adult participants are more probable to approach the problem in an analytic way, but still, we thought that learning the word-pairs first could help them recognizing the centre-embedded structure of sentences. For this reason, in this experiment, we used *staged input* with *starting small*: participants first listened to two-word-long sentences (Stage 1), then four-word-long sentences, and finally six-word-long sentences.

Computational models (e.g., Elman, 1993) and AGL experiments (e.g., Cochran, McDonald, & Parault, 1999) showed that starting small can be essential for learning complex grammars; however, there are both simulations and empirical works with opposite conclusions (e.g., D. L. T. Rohde & Plaut, 1999; Douglas L. T. Rohde & Plaut, 2002). Moreover, Conway, Ellefson and Christiansen (2003) found that starting small helped learning CER when the task was visually presented, but not when it was presented auditorily. In their experiment there were six words in both classes and each word had exactly three possible pairs from the other word class. We assumed that the starting small paradigm is more probable to help when word-pairing is one-to-one, i.e., it is easier to memorize. However, with the present experiment we did not want to decide whether starting small and staged input helps learning CER, this learning paradigm was simply one of the modifications we administered (compared to the previous experiment), because we thought it would make the task easier. During the two-word-long stage participants supposedly memorize word-pairs and familiarize with the idea that certain words belong together. To help this process word-pairs started with the same consonant, e.g. de-do, ti-tu.

We also changed the procedure: while in the previous experiment the only opportunity for learning was during the short habituation period, in this experiment learning occurred through three types of tasks. Apart from the listening task which was similar to the habituation of the previous experiment more active hypothesis searching was encouraged by two other tasks: discrimination of grammatical and ungrammatical sentences and sentence completion. During these tasks feedback was given so that participants could experiment with different hypothesis, try different rules and see whether they are valid or not. These new tasks also made it possible to examine the comprehension-production asymmetry known to be present during language development (Hendriks & Koster, 2010). The discrimination task provides a measure for comprehension, while sentence completion models language production because it requires active generation of grammatical sentences.

All these changes were introduced to help participants learn the rule. However, the main purpose of this experiment was to compare two conditions: for half of the participants sentences were presented auditorily only, while for the rest of the participants auditory strings were complemented by a written script of all of the tasks. This condition changes the task in two important ways. First of all, it decreases working memory load, since participants do not have to remember the sentences to be able to parse them which supposedly makes the task easier. Second, while auditory sentences are inherently sequential in time, written sentences can be seen as a whole which makes it possible to go back and forth between words easily while looking for regularities. Since regularities in CER involve long-range dependencies between words, a more global view of sentences might help recognizing them. We predicted that participants in both conditions will be able to learn CER, but learning would be faster when the written script is provided.

5.2 Materials and Methods

5.2.1 Participants

Fourteen secondary school students participated in this study. They were randomly assigned to two equal groups. The Reading group consisted of 4 females and 3 males, while the Listening group consisted of 2 females and 5 males. One participant was 19 years old, another 17 years old; all other participants were 18 years old.

One male participant from the Listening group opted out from the experiment halfway through, and thus was excluded from further analysis.

5.2.2 Stimuli

The vocabulary (Table 7) consisted of 16 consonant-vowel syllables arranged in two classes and eight pairs. Words in Class *A* always ended in *e* or *i*, while words in Class *B* always ended in *o* or *u*. Each word from Class *A* had exactly one pair from Class *B* and vice versa. Word-pairs started with the same consonant.

Grammatically correct sentences were composed from the vocabulary according to the centre-embedded rule (A_1B_1, $A_1A_2B_2B_1$, $A_1A_2A_3B_3B_2B_1$). In the case of two-word-long sentences, grammatically incorrect sentences were generated from non-pair words from Class *A* and *B*, always keeping the correct order of classes – in this way only the consonants (indicating pair membership) were incorrect not the vowels (indicating class membership). Longer grammatically incorrect sentences were generated from grammatically correct sentences by swapping words in the second half of the sentence: in the case of four-word-long sentences we swapped the third and fourth words, while in the case of six-word-long sentences we swapped the fourth and the fifth word.

Table 7. Vocabulary

Class	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6	Pair 7	Pair 8
A	de	gi	ke	mi	ne	ri	se	ti
B	do	go	ku	mu	no	ro	su	tu

5.2.3 Procedure

Participants were presented with eight blocks of training and testing. Each block consisted of five tasks. The first task was listening during which participants simply listened to the sentences. The second task was discrimination. In this task participants listened to randomly ordered grammatically correct and incorrect sentences and had to decide whether they conformed to the rule or not. Participants indicated their answers by circling Yes/No on the score sheet. After each sentence participants received the correct answer and were asked to self-correct their answers. The third task was sentence completion. In this case, sentences were correct but one word was missing from the second half of the sentences. The missing word was indicated by an "sss" (hissing) sound. Participants had to write down the missing word, after which the correct answer was given. The fourth and fifth tasks were discrimination and sentence completion again but this time without feedback. Our reason to provide the same tasks within blocks with and without feedback was that we wanted to control for possible cheating on the part of participants. Even without the intent to cheat it is possible that self-corrections after feedback are not discriminable from other self-corrections.

The first task (listening) consisted of 10 sentences, and all the other tasks consisted of 5 sentences. In this way, each block contained 10 + 4*5 = 30 sentences. The 8 blocks of training and testing involved three stages according to the length of the sentences: The first two blocks trained and tested two-word-long sentences, the next three blocks presented four-word-long sentences and the last three blocks presented six-word-long sentences. All in all, the whole procedure contained 240 sentences and lasted about half an hour.

Participants were sitting in a classroom, 13 of them at the same time. One of the participants was tested on a different occasion. The experimenter read the instruction to participants, explaining the length and type of different tasks. The instructions about the grammar were the following: "You will hear the sentences of an artificial language. The sentences are composed of two-letter monosyllabic words according to a rule. Your task is to find out the rule."[12]

[12] In Hungarian: *A mesterséges nyelv értelmetlen, egy szótagos, két betűs szavakból áll, amikből valamilyen szabály alapján mondatokat szerkesztettem. A feladat az lesz, hogy próbáljátok kitalálni, milyen szabály alapján szerkesztettem a mondatokat.*

After the instructions all participants received their score sheets (see Appendix). For the Reading group, the score sheet included the script of the whole session, i.e., each sentence was written down. For the sentence completion, a gap indicated the missing words. For the Listening group, only the type of tasks was indicated and space was provided for putting down the answers. Participants were asked not to write down the sentences they hear.

Sentences were generated by a text to speech software (SpeechPad with Infovox desktop v2.210, publisher: Acapela group), using a male voice (Vittorio 22k_ID2210). Length of words were approximately 300 ms, length of silent gaps between words were about 400 ms. There was a 1200 ms silent interval between sentences during the listening task. During all the other tasks 3 sec were provided for the participants to record their answers before the next sentence or the feedback followed.

After 8 blocks (about 30 minutes) of training, participants were asked to write down the rule they deduced from the tasks. Finally, there was a perceptual task: participants heard all the words in the vocabulary in a random order and were asked to write them down to test if they could hear the sounds clearly.

5.3 Results

5.3.1 Perceptual task

Average number of correct words was 15.4 (SD = 1.0). Nine out of 13 participants managed to write down all the words correctly. Four participants (all of them from the Listening group) had 1 to 3 errors; there were eight misspelled words in total. Misspelled words were noted and were later used for the scoring of the sentence completion tasks.

5.3.2 Scoring

The maximum score in each task (discrimination with and without feedback and sentence completion with and without feedback) was 5 for each participant (1 score / sentence). For the discrimination tasks, we simply counted the number of correct answers in each task. The scoring of the sentence completion tasks was a little bit more complicated. For the four participants who had errors in the perceptual task we checked whether the misspelled words consequently reappeared in the sentence completion tasks. If the misspelled word was misspelled at all occurrences during the tasks, we scored them as correct, because this means that all of these errors were due to misperceiving the word. Four out of eight misspelled words were like this. In two cases, misspelled words never reappeared during the tasks and were spelled correctly throughout, which means that misperceiving them during the perceptual task was just by chance and misperception did not influence the tasks.

For all participants, if the missing word given by the participant was incorrect, but started with the correct consonant (i.e. the pairing was correctly identified) and ended with *a* instead of *o* or *o* instead of *a* half score was given. This means that the centre-embedded rule and the rule for pairing (pairs start with the same consonant) and grouping words (words in Class *A* end in *e/i*, words in Class *B* end in *o/a*) was already understood, only the exact words had not been memorized yet.

An independent colleague analysed participants' written formulation of the rule. Answers were regarded as correct if they expressed somehow the centre-embedded structure of sentences. Most correct answers included the words "symmetric", "mirrored", an explicit formula of the sentences (e.g., abcc'b'a' or 123321), or the explanation of the structure of the sentence (e.g., "If you start form the middle of the sentence and go in both directions words start with the same consonant…").

5.3.3 Discrimination and sentence completion tasks, formulation of the rule

5.3.3.1 Is there an effect of written sentences?

There was a striking difference between the performances of the two groups of participants. The Reading group scored maximum in almost all of the tasks, from the very beginning: there were only 2.5 errors in total (out of 32 tasks, 5 sentences each). All of them correctly formulated the rule too. However, the Listening group's performance was always below 90% with the exception of one task (the discrimination task without feedback during the second block).

We compared the performance of the groups with *t* tests. In those tasks where the Reading group reached maximum score, we assumed homogeneity and compared the scores of the Listening group with the hypothetical value of 5 with one-sample *t* test. In those three tasks where the Reading group did not reach the maximum score (discrimination without feedback in Block 5, discrimination with feedback and completion without feedback in Block 7) we compared the groups with independent samples *t* test.

The difference between groups was significant in 27 of the 32 tasks (see *p* and *t* values in

Table 8). All tasks where the difference was not significant were at Level 1 of the training, where participants had to learn word-pairs, including all of the discrimination tasks in Block 1 and 2, and the completion task without feedback in Block 2. This is quite surprising. We expected a difference between groups at Level 1 since this is the stage where the Listening group has to memorize all word-pairs; whereas the Reading group has everything written down (they could look back on the script of the listening task when doing the discrimination and completion tasks of the same block). The Reading group made no mistake at this level which means that the task was very easy for them indeed; nevertheless the Listening group also achieved high performance, especially in discrimination tasks. On completion tasks they performed worse than the Reading group (except for the last completion task on this level), a pattern that shows the comprehension-production asymmetry.

Table 8. Results of the *t* tests comparing the Reading group and the Listening group in each task. Disc.F. = Discrimination task with feedback, Comp.F. = Sentence completion task with feedback, Disc. = Discrimination task without feedback, Comp. = Sentence completion task without feedback.

Level:	Level 1							
Block	Block 1				Block 2			
Task:	Disc.F	Comp.F	Disc.	Comp	Disc.F.	Comp.F	Disc.	Comp.
t	-1.464	-5.398	-	-3.093	-1.464	-2.939	-	-1.908
p	**.102**	.002	**.051**	.014	**.102**	.016	**.182**	**.058**

Level:	Level 2											
Block	Block 3				Block 4				Block 5			
Task:	Disc.F	Comp.F	Disc.	Comp	Disc.F.	Comp.F	Disc.	Comp.	Disc.F	Comp.F	Disc.	Comp
t	-2.666	-3.322	-	-5.196	-3.953	-3.322	-	-2.937	-2.169	-2.395	-	-2.902
p	.023	.011	.019	.002	.006	.011	.019	.016	.041	.031	.028	.017

Level:	Level 3											
Block	Block 6				Block 7				Block 8			
Task:	Disc.F	Comp.F	Disc.	Comp	Disc.F.	Comp.F	Disc.	Comp.	Disc.F	Comp.F	Disc.	Comp
t	-2.150	-2.500	-	-3.800	-4.000	-3.796	-	-2.454	-2.712	-4.332	-	-6.220
p	.042	.027	.003	.007	.004	.007	.017	.029	.021	.004	.009	.001

*Unpaired *t* test; all other values are from one-sample *t* tests (see text). df = 5 in each case, *p* values are one-sided.

The difference was significant between groups in all the other tasks on the following levels, which meant that not having to use working memory and being able to look at the sentences as a whole made the tasks easier on Level 2 and 3. In other words, recognizing the centre-embedded rule is easier when participants can see the whole sentence all at once. The main difference between listening to a sentence and reading a sentence is twofold: first of all, listening is linear: one hears the words as a sequence. Reading is more global, one can look at the sentence all at once, can go back and forth between words. When learning CER participants have to match words from the end of the sentence with words from the beginning of sentence which can be done directly with the help of the script, while during listening, working memory plays a big role. All in all, it is obvious that the computational load is much less when reading vs. when listening to a sentence with CER. These results are in accord with the fact that CER is more frequent in written than in spoken language (Karlsson, 2007).

5.3.3.2 Did participants learn CER?

First we performed one-sample *t* tests to see whether participants as a group performed better than chance in the discrimination tasks (test value = 2.5). It is not possible to calculate *t* when *SD* = 0, so for the Reading group we only performed the test when the group was not homogeneous (where there was a mistake). In these tasks (discrimination without feedback in Block 5 and with feedback on Block 7) the Reading group performed better than chance, $t(6) = 16.5$, $p = .000$. For the statistics of the Listening group see

However, these are the results of the discrimination tasks only which represent passive rather than active knowledge. Completion tasks test active knowledge, but it is not possible to determine the test value in these tasks, so we did not perform *t* tests for these. But if we compare the scores for discrimination and completion tasks it is clear that the classic production-comprehension asymmetry shows here. Figure 10 highlights the differences between the task types in the Listening group: the scores for the discrimination tasks are higher than the scores for the sentence completion tasks in all but one case (in Block 7). A repeated measures ANOVA with within subject factors of block (with 8 levels), feedback (with 2 levels: with feedback and without feedback) and type of task (with 2 levels: discrimination and completion) showed that the type of task has a significant effect on performance, $F(1, 5) = 22.491$, $p = .005$.

Table 9. Participants performed better than chance in only 5 of the tasks. Four of these were in Block 1 and 2, i.e., on Level 1 of learning; the fifth was in Block 7. From this it seems that the Listening group successfully learned the word-pairs but failed to recognize the centre-embedded rule.

However, these are the results of the discrimination tasks only which represent passive rather than active knowledge. Completion tasks test active knowledge, but it is

not possible to determine the test value in these tasks, so we did not perform *t* tests for these. But if we compare the scores for discrimination and completion tasks it is clear that the classic production-comprehension asymmetry shows here. Figure 10 highlights the differences between the task types in the Listening group: the scores for the discrimination tasks are higher than the scores for the sentence completion tasks in all but one case (in Block 7). A repeated measures ANOVA with within subject factors of block (with 8 levels), feedback (with 2 levels: with feedback and without feedback) and type of task (with 2 levels: discrimination and completion) showed that the type of task has a significant effect on performance, $F(1, 5) = 22.491$, $p = .005$.

Table 9. Results of the *t* tests for the discrimination tasks of the Listening group. Df = 5 in each case. Significant values are in bold.

Block	Discrimination tasks	t	p
Block 1	with feedback	5.855	**.002**
	without feedback	4.152	**.009**
Block 2	with feedback	5.855	**.002**
	without feedback	14.000	**.000**
Block 3	with feedback	1.777	.136
	without feedback	.000	1.000
Block 4	with feedback	1.976	.105
	without feedback	1.019	.355
Block 5	with feedback	1.898	.116
	without feedback	1.356	.233
Block 6	with feedback	2.457	.057
	without feedback	.698	.516
Block 7	with feedback	-.271	.797
	without feedback	3.322	**.021**
Block 8	with feedback	1.356	.233
	without feedback	2.335	.067

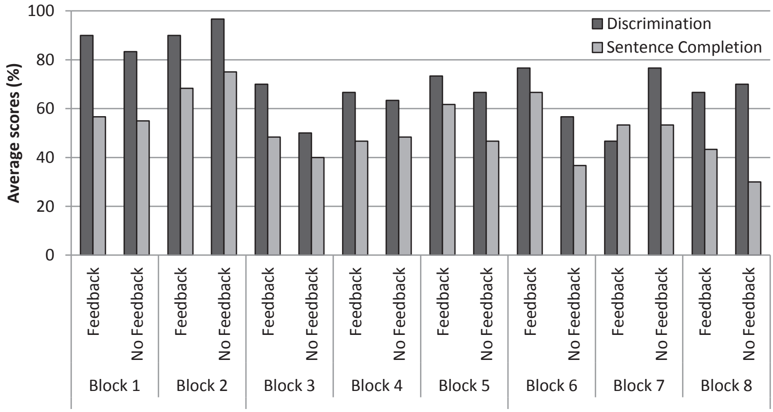


Figure 10. Average performance of the Listening group on all tasks.

Second, an independent (blinded) colleague analysed the participants' written formulation of the rule. As we mentioned earlier, all participants in the Reading group, but only half of the participants (three out of six) described CER correctly.

Third, we wanted to examine the performance of participants on an individual level. For this, we merged the scores from the tasks with and without feedback for both task types in each block. Originally, we planned discrimination and completion tasks both with and without feedback for a practical reason: we were not sure that participants would be able to resist correcting their mistakes when feedback was given. But it seemed that participants were not interested in doing this, and observing the test sheets also confirmed that participants clearly indicated their mistakes (they were asked to write the correct answer after the feedback next to their original answer and to cross their original if it was incorrect). To make sure that participants did not cheat we performed a repeated measures ANOVA (see above) with within subject factors of block (with 8 levels), feedback (with 2 levels: with feedback and without feedback) and type of task (with 2 levels: discrimination and completion). Feedback had no effect, $F(1, 5) = 2.074$, $p = .145$, nor did task type × feedback, $F(1, 5) = 1.909$, $p = .226$, which means that participants did not cheat when feedback was given. This enabled us to merge the scores from the tasks with feedback with those without feedback in each block. In this way we gained two scores for each block, one from the discrimination tasks and the other from the completion tasks, each with a maximum score of 10 (see Table 10).

According to the criterion that is usually used in the literature (e.g., Bahlmann, Schubotz, & Friederici, 2008) 9 or 10 scores in two consecutive blocks is considered as successful learning of the task of a given level (this is significantly greater than chance according to the binomial test, p = .01074, in the case of the discrimination task). We analysed the two types of tasks separately in this regard.

For discrimination half of the participants in the Listening group successfully learned the word-pairs (#1, 3, 4), and two more participants were close to it (#5 had 9 + 8 scores, #6 had 8 + 10 scores). In other words, there was only one participant out of six, who did not get close to learning the task (#2). On the contrary, there was only one participant who passed the criterion for completion tasks on this level (#3) and one more who got close to it (#1 with 8 + 9.5 scores). On the following levels performance dropped. Those participants who could not formulate the rule (during the last task at the end of the training) did not get close to the criterion either in any of the tasks. Three participants described the rule correctly. One of them mastered both tasks on both levels (#4). Another participant (#3) passed both tasks on Level 2, but none of them on Level 3 (although he missed the criteria for discrimination by only 1 score). The last participant who correctly described the rule did not pass the criterion on any of the levels or tasks; however, she got close to it on Level 3: scored 9 + 6 + 9 on discrimination tasks and 8 + 9 + 7 on completion tasks. All in all, it seems that these three participants learned CER because they could formulate the rule, but two of them had problems during applying the rule, especially in the case of six-word-long sentences, when the working memory load is higher.

Table 10. Merged (with and without feedback) scores of participants in the Listening group. D = discrimination task, C = completion task; Scores higher than the criteria are highlighted.

P #	Level 1				Level 2						Level 3						Rule
	Block 1		Block 2		Block 3		Block 4		Block 5		Block 6		Block 7		Block 8		
	D	C	D	C	D	C	D	C	D	C	D	C	D	C	D	C	
1	10	8	10	9.5	4	2	5	4.5	7	3	7	3	4	3	5	2	0
2	6	5	8	5.5	3	0	4	1	6	2.5	5	1	6	3	4	2	0
3	10	9	10	9.5	10	7.5	9	9	10	10	8	8	8	6	9	4	1
4	9	4	10	8	10	9	10	9	10	9	7	9	9	10	9	7	1
5	9	2.5	8	2.5	2	3	5	0	2	0	4	2	4	1	5	0	0
6	8	5	10	8	7	5	6	5	7	8	9	8	6	9	9	7	1

5.4 Conclusions

Two groups of participants were tested on discrimination and sentence completion tasks on three increasing levels of CER. Both groups of participants listened to the sentences, but the Reading group received all the stimuli written down, while the Listening group did not receive additional information beyond the auditory stimuli. There are different computational requirements on each level of learning: on the first level (two-word long sentences) word-pairs have to be memorized; on the second (four-word long sentences) participants have to recognize CER and on the third (six-word long sentences), they have to generalize the rule and apply it to longer sentences. Discrimination and sentence completion tasks were supposed to model language comprehension and production.

Our predictions were verified: (1) Learning CER (on Level 2 and 3) is easier when the written sentences are available to the learner which is congruent with the fact that CER is more frequent in written than in spoken language. (2) Discrimination tasks are easier to master than completion tasks which reflect the classic comprehension/production asymmetry. All participants in the Reading group and half of the participants in the Listening group recognized CER in the sentences based on their written formulation of the rule. Surprisingly, two participants from the Listening group who mastered discrimination tasks on Level 1 could not learn CER on the following levels. Also, group performance in the Listening group was above chance on Level 1 discrimination tasks, but not on higher level discrimination tasks and the difference was not significant between groups on Level 1 discrimination tasks. We expected that memorizing word-pairs would be the hardest task, and once it is done, recognizing CER would be relatively easier, since the word-pairs were absolutely new and artificial for participants whereas CER as a structure is present in all natural languages. On the contrary, it seems that memorizing word-pairs was not the bottleneck, even though for this task only two blocks were provided, whereas for the higher levels there were three blocks; it was Level 2 and 3 that were harder to master.

6 Recursion and semantics

This chapter is based on a published paper of Fedor, Varga, & Szathmáry (2012). The original title was: *Semantics boosts syntax in artificial grammar learning tasks with recursion.* In the first two experiments (Chapter 4 and 5) we found that CER is harder to learn than it was expected. Granted, when participants did not have to rely on their working memory (the Reading group in Chapter 5), it was an easy task. Although CER is less frequent in spoken than in written language humans can cope with centre-embedded sentences in spoken language too without the help of a written script. We hypothesised that the presence of semantics in natural language could trigger parsing CER, which would explain why CER is so difficult in AGL tasks where semantics is absent.

6.1 Abstract

Centre-embedded recursion (CER) in natural language is exemplified by sentences like *The malt that the rat ate lay in the house*. Parsing centre-embedded structures is in the focus of attention because this could be one of the cognitive capacities that make humans distinct from all other animals. The ability to parse CER is usually tested by means of artificial grammar learning (AGL) tasks, during which participants have to infer the rule from a set of artificial sentences. One of the surprising results of previous AGL experiments is that learning CER is not as easy as we thought. Because artificial sentences lack semantic content, we hypothesized that semantics could help humans learn the syntax of centre-embedded sentences. To test this, we composed sentences from four vocabularies of different degrees of semantic content due to three factors (familiarity, meaning of words and semantic relationship between words). According to our results, these factors have no effect one-by-one but combined they make learning significantly faster. This leads to the assumption that there were different mechanisms at work when parsing CER in natural and in artificial languages. This finding questions the suitability of AGL tasks with artificial vocabularies for studying the learning and processing of linguistic centre-embedded recursion.

6.2 Introduction

Artificial grammar learning (AGL) tasks are widely used to test the abilities of different species in learning different grammatical rules. During these tasks participants are usually trained and tested on a set of artificial sentences to assess whether they could master the grammatical rule underlying these sentences. Sentences are composed of a set of nonsense "words" (the vocabulary), which could be anything from actual letters to geometrical shapes but are usually consonant-vowel (CV) syllables. The theory behind this paradigm is that removing semantics from a language makes it possible to research its pure syntax.

After the influential paper of Hauser et al. (Hauser, et al., 2002), active research started using AGL tasks investigating a particular grammar, called centre-embedded recursion (CER). In natural language, CER is exemplified by sentences like "*The malt that the rat ate lay in the house*". There are three main characteristics of this sentence: (1) A phrase (*that the rat ate*) is embedded within another (*the malt lay in the house*); (2) there are within-phrase dependencies between different classes of words (here, nouns and verbs) the *malt* – *lay* and the *rat* – *ate*; and (3) there is also a dependency between phrases: *the rat ate* qualifies *the malt* (only the malt that the rat ate would do, not just any malt).

The first generation of AGL experiments on CER (e.g., Gentner, et al., 2006) conformed only to the first characteristics of CER. In these experiments, a four-word-long sentence could be described by the formula of *AABB* (A^nB^n in general), where *A*s and *B*s are arbitrary words from two distinct classes of artificial words. It means that *AB* phrases are embedded within each other but the dependencies between and within phrases are not modelled. Due to these simplifications, it was possible to solve the tasks (discrimination between grammatical and ungrammatical sentences) without recognizing the recursive structure of sentences, by simply matching the number of *A*s and *B*s (Corballis, 2007a, 2007b; Perruchet & Rey, 2005a).

A second generation of experiments tried to get around this problem, by establishing *A*-*B* word-pairs. These sentences could be described by the formula $A_1A_2B_2B_1$, where indices denote dependencies between *A*s and *B*s; the between-phrase dependencies are still missing but the within phrase dependencies are present. These experiments yielded various results. In the experiment of Perruchet and Rey (2005a) human participants were not able to learn the grammar after 3 minutes of habituation. Similarly, in de Vries at al.'s study (2008) 50 minutes of alternating familiarization and test blocks with feedback (230 sentences in sum) were not enough for participants to recognize the structure of sentences. However, there were two studies, in which participants managed to learn CER: that of Bahlmann et al. (2008) and Lai and Poletiek (2011). These studies share a number of methodological points (one or more of which were missing from the previous studies): word-pairs and word classes (*A* and *B*) were distinguished by phonological cues; training was staged with sentences of increasing length (starting small paradigm); there were alternating habituation and testing blocks and during testing feedback was provided.

These experiments require different computations from learners at different training stages. The starting small paradigm with staged input means that learners are first exposed to shorter and then increasingly long sentences. In the case of CER the first stage of learning involves two-word-long sentences (i.e., word-pairs); the second stage involves four-word-long sentences; and the third stage involves six-word-long sentences. On the first stage, where two-word-long sentences are presented, associative learning is required to memorize word-pairs (which is supposedly helped by phonological cues). On the next stage, the task is to recognize the centre-embedded structure of sentences. Because of the feedback, participants probably engage in active rule searching as opposed to passive incidental learning. The last stage tests generalization of the rule to longer sentences. Usually it is obvious from the instructions given to participants that the rule is the same throughout, so at this stage participants have to learn how to apply the previously learnt rule effectively.

Even in these experiments, where learning was successful, extensive training was needed to reach the desired performance. This is quite surprising, seeing that CER is present in all known human languages (although, see Everett, 2005) and the ability to parse it was supposed to be a natural and straightforward human ability. Simplifying natural language to syntax + semantics and comparing it to artificial languages that lack the latter leads to the idea that it is indeed the absence of semantics that makes it so difficult to recognize the centre-embedded structure in artificial sentences. We designed an experiment to test this hypothesis, in which we trained participants on artificial sentences that involved different degrees of semanticity. We predicted that artificial sentences with semantic content make learning easier at all stages compared to sentences with no semantic content.

6.3 Materials and methods

6.3.1 Participants

67 Hungarian native speaker participants (two participants were bilingual), mainly university students participated in this study (M = 22.1 years, SD = 3.8; 30 female and 37 male). They were randomly assigned to four groups: there were 18 participants in Group WS, 16 participants in Group WR, 16 participants in Group NR1 and 17 participants in Group NR2. Groups were named after the vocabulary types they were trained on (see below).

They had no known disorder or had not taken any drugs that might have influenced memory or attentional abilities. They had normal, or corrected to normal vision. Participants received course credit, or small refreshments (chocolate or beer) for their participation.

6.3.2 Stimuli

Vocabularies from which sentences were composed of contained six pairs of words; six words comprised in Class *A* and six words comprised in Class *B*. Every word had exactly one pair from the other class. There were four distinct vocabularies, one for each group of participants (Table 11). The first vocabulary consisted of two-letter Hungarian words that were selected during a previous short study. In this study, participants (different from those in the present experiment) had to make pairs from a pool of 21 Hungarian two-letter nouns based on free association. Those pairs that were chosen most often made up the first vocabulary. As it can be seen pairing was mainly based on the semantic relationship of the words, that is why we labelled this vocabulary WS (Words with Semantic relatedness). The second vocabulary consisted partly of words from Vocabulary WS: Class *A* was the same as in Vocabulary WS, but Class *B* contained different words that were chosen in a way that there were no semantic relatedness between *A* and *B* words. Moreover, we chose words that had one letter in common with a Class *B* word in Vocabulary WS; hence the two vocabularies were phonologically as similar as possible. We labelled this vocabulary WR (Words Randomly paired).

Table 11. Four different vocabularies from which sentences of the artificial language were generated for the four distinct groups of participants.

Vocabulary WS		Vocabulary WR		Vocabulary NR1		Vocabulary NR2	
Class *A*	Class *B*	Class *A*	Class *B*	Class *A*	Class *B*	Class *A*	Class *B*
eb (dog)	ól (kennel)	eb (dog)	ón (tin)	ev	ób	nu	zi
én (me)	te (you)	én (me)	tó (lake)	éz	ta	gi	pe
év (year)	ősz (autumn)	év (year)	ős (ancestor)	őgy	fe	ru	ve
fű (grass)	fa (tree)	fű (grass)	ma (today)	fé	ísz	fe	ko
íny (gum)	íz (flavour)	íny (gum)	ív (arc)	ít	őn	bi	mo
kő (stone)	út (road)	kő (stone)	úr (gentleman)	kű	úl	lu	co

We chose these vocabularies to test whether semantic relationship between words has an effect on learning. We could have generated Vocabulary WR from the words of Vocabulary WS by randomizing the pairs; however, it would have resulted in a vocabulary where there are obviously related words that are not treated as pairs, which could have made the task more difficult. Therefore we composed Vocabulary WR partly from Vocabulary WS (Class *A*) and partly form new words (Class *B*) so that there is no obvious semantic relationship between any two words.

The third and fourth vocabularies contained non-words (consonant-vowel syllables) randomly paired, so we labelled them NR1 and NR2 (Non-words Randomly paired). Vocabulary NR1 was generated mainly from the letters of words in Vocabulary WS in such a way that no word had a meaning, not even reading backwards (we had to change some of the letters to meet this criterion). Care was taken that words had no meaning in most other languages that Hungarian students usually learn and that word-pairs (read together as one word) did not make sense either. We kept the position of letters in words as in Vocabulary WS, as much as possible. In this way, this vocabulary was phonologically similar to Vocabulary WS and WR, but the words had no meaning. Lastly, Vocabulary NR2 consisted of non-words that were similar to vocabularies of other studies that were conducted with German speaking participants (e.g., Bahlmann, et al., 2008; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006). There were no long vowels in this vocabulary, which are very common in Hungarian words.

Our motivation to test participants on two different non-word vocabularies was that we realized that Hungarian students learned much slower in our previous study (Fedor & Szathmáry, 2009), than German students in Bahlmann et al.'s study (2008), however, the circumstances were quite similar. We had the feeling that the vocabulary that was used in both studies could be more "familiar" to German native speakers than to

Hungarians (even though the vocabulary was phonotactically legal in Hungarian too). To test this effect, we constructed Vocabulary NR1 using Hungarian-specific vowels, and thus it sounded more "Hungarian-like" than Vocabulary NR2.

Sentences composed from these vocabularies represent four different levels of diversion from natural language (Table 12) according to three criteria: phonetic familiarity, words with meanings and semantic associations between words. Vocabulary NR2 is the least natural; it does not meet any of the above mentioned criteria. All the other vocabularies sound familiar to Hungarian participants. Vocabulary NR1 is composed of non-words that have no meaning, whereas the remaining two vocabularies are composed of natural words with meaning. Only Vocabulary WS meets all three criteria, however there are still a lot of differences with natural language.

Table 12. Similarity of vocabularies to natural language according to three criteria.

Vocabulary:	WS	WR	NR1	NR2
Does the vocabulary sound phonetically familiar?	Yes	Yes	Yes	No
Do the items in the vocabulary have meaning?	Yes	Yes	No	No
Is there semantic relationship between items?	Yes	No	No	No

The rule of centre-embedded recursion was used to compose sentences from these vocabularies. In case of two-, four- and six-word-long sentences the rules were A_1B_1, $A_1A_2B_2B_1$ and $A_1A_2A_3B_3B_2B_1$, respectively. Indices denote dependencies between words, i.e. an *A* word and a *B* word with the same index make up a word-pair. In this way 6 two-word-long, 30 four-word-long and 120 six-word-long grammatical sentences were composed with each vocabulary.

Ungrammatical sentences were generated by randomly replacing one of the words in the second half of a grammatical sentence by another *B* word, thus violating the structure of word-pairs, but not the structure of word-classes (*A*s and *B*s) in sentences, thus ensuring that the error was detectable provided that one was aware of the centre-embedded structure of word-pairs. *B* words that were already in the sentence were not excluded from being replacements, thus word repetitions could occur in four-word-long and six-word-long ungrammatical sentences. This decision was made in accordance with Bahlmann et al.'s study (2008), where such repetitions were also allowed, because we wanted to compare the performance of our participants on Vocabulary NR2 with the performance of participants in the above mentioned study[13]. Replacements were performed in all possible positions (but only in one position in a sentence), thus in the second position of two-word-long sentences, in the third or fourth position of four-word-long sentences and in the fourth, fifth or sixth positions of six-word-long sentences.[14]

6.3.3 Procedure

The procedure followed the schema of the learning period of the experiment of Bahlmann et al. (2008). In the beginning of the training, participants were given the instructions that they would read the sentences of an artificial language, and their task was to find out the rule according to which the sentences were composed. Training of participants was performed according to the "starting small" paradigm with staged input (Conway, et al., 2003): started with two-word-long (Level 1) followed by four-word-long (Level 2) and then six-word-long sentences (Level 3).

A training block consisted of a set of ten familiarization sentences and a set of ten test sentences. The familiarization set started with an instructional sentence (the whole sentence presented all at once): "Please read carefully the following sentences corresponding to the rule!" During familiarization, sentences followed each other separated only by a fixation cross in the middle of the screen. All sentences were grammatical. Test sets were also anticipated by an instructional sentence: "Please decide, whether the following sentences correspond to the rule or not!" Test sets were compiled from five grammatical and five ungrammatical sentences, randomly ordered. There was a fixation cross before and a choice of "Yes" or "No" after each sentence. Participants had 3 seconds to answer and then feedback was given: for 250 ms the right answer flashed on the screen.

Familiarization and test sentences were randomly chosen from the pool of grammatical and ungrammatical sentences without replacement until all sentences were

[13] It can be argued that repetitions make it possible to detect ungrammaticality without learning the grammar of sentences, however it is very unlikely that participants could pass the test if their decisions had been based solely on repetitions (see calculations for this probability in the results).

[14] As an example, see supporting online material for the entire pool of grammatical and ungrammatical sentences for Vocabulary WS, from which training and test sentences were randomly chosen for each participant in Group WS.

used. After that all sentences were placed back to the pool and the same procedure was applied again.

Sentences were visually presented on a computer screen with one word at a time. The first word of sentences started with a capital letter and sentences were closed by a full-stop. Words were showed for 800 ms followed by a 200 ms gap. The fixation cross was showed for 1000 ms before every sentence.

If a participant had reached nine or ten correct answers in two consecutive training blocks, then the next level with longer sentences followed. Each level consisted of as many blocks as the participant needed to reach the required performance. If a participant had not mastered a level during 20 blocks, the test was finished without proceeding to the higher levels.

After the test was finished, participants were asked to write down the rule that they deduced from the sentences.

6.4 Results

To find out whether the difficulty of the task was different in the four groups, we performed two kinds of analysis. First we compared the success rate of participants in the four groups (whether they reached the required performance on the different levels; and the correctness of their written formulation of the rule), and then we compared the number of training blocks they needed to finish the training.

Whether passing the 90% performance criterion means that the participant understands the rule can be questioned. Because there is a relatively low number of grammatical sentences in Level 1 (6 sentences) and Level 2 (30 sentences), sentences could be memorized instead of learning the rule (in fact, sentences – word-pairs – had to be memorized in Level 1). However, participants who memorized four-word-long sentences without understanding the rule would not be able to pass the criterion on six-word-long sentences (unless they memorized six-word-long sentences too, which is unlikely). Because there was no participant who passed Level 2 but did not pass Level 3 we can exclude this possibility.

A participant could have passed the 90% performance criterion basing his decisions solely on detecting word repetition in ungrammatical sentences if there had been 4 or 5 ungrammatical sentences with word repetition in two consecutive blocks. This means 8-10 sentences with word repetition in sum out of 10 ungrammatical sentences in two consecutive blocks: if a participant categorizes sentences with repetition as ungrammatical and sentences without repetition as grammatical, he could have 18-20 correct answers in two blocks and could pass the test. Obviously this is undesirable, because we do not want to confound this simple strategy with true understanding of the grammar. However, we did not worry about this, because its probability according to the binomial distribution is very small: it is $3.5006*10^{-5}$ and 0.0202 in the case of four-, and six-word long sentences, respectively (calculated from the average percentage of ungrammatical sentences with word repetition across vocabularies: 18% and 43%). In fact, we checked the last two blocks in Level 3 of successful participants, and we found only 3 cases where more than 7 ungrammatical sentences occurred with word repetition. None of these participants mentioned word-repetition in their written formulation of the rule. There were only one participant in the four groups that mentioned that sentences with word repetitions were not correct, but he was not successful in passing Level 2.

There was one participant in Group WR and one in Group NR2 who did not learn the word-pairs and thus was excluded from all further analyses. All other participants reached the 90% criterion on word-pairs (Level 1) and proceeded to Level 2. Two participants in Group NR1 and six participants in Group NR2 did not learn the recursive rule in four-word-long sentences during the 20 training blocks provided (400 sentences) and thus did not proceed to Level 3. All successful participants on Level 2 were able to reach the 90% criterion on Level 3 too. According to the Chi-square test, the success rate of participants on Level 2 and their group membership were related, $\chi^2(3, N = 65) = 14.04$, $p = .003$, which implies that the success rate (which was influenced by the difficulty of the task) was significantly different in the four groups. Note that this difference results only from participants' performance on Level 2.

An independent colleague analysed participants' written formulation of the rule. Answers were regarded as correct if they expressed somehow the centre-embedded structure of sentences. Most correct answers included the words "symmetrical", "mirrored", "embedded", or an explicit formula of the sentences (e.g., "abccba" or "123321"). The overlap was not perfect between success according to the 90% criterion and correctness of the written rule: 8 participants who were successful according to the 90% criterion were unable to write down the rule (3 from Group WS, 1 from Group WR, 1 form Group NR1, and 3 from Group NR2). While it can be a far reaching question what these participants really learnt, the true understanding of the rule by those participants who passed both criteria cannot be questioned. According to the Chi-square test, the success rate of participants on the formulation of the rule and their group membership were related, $\chi^2(3, N = 65) = 12.143$, $p = .007$, which enforces the previous finding.

For comparing the number of training blocks needed in the four groups we included the data of unsuccessful participants (i.e. we used 20 blocks as their measure of performance on Level 2 in the analysis), noting that we do not know the accurate number of training blocks they would have needed to reach the criterion on Level 2; the only thing we know is that it would be more than 20. Fortunately, this decision did not affect our statistics (see Footnote 3). Also, we note that the number of training blocks to reach criterion on Level 3 is missing from the analysis for these participants.

The average number of blocks needed to finish all three levels in Group WS was 7.28 (SD = 3.03). Most of the participants needed only 2 blocks/level (note that this is the least possible according to the training regime), which means that their performance was 90% or above after reading only 10 sentences. Group WR needed 12.27 blocks (SD = 3.788), Group NR1 16.94 blocks (SD = 5.260) and Group NR2 20.25 blocks (SD = 7.646) to finish all levels on average, and the difference was significant between each pair of groups except for Group NR1 and NR2 [Kruskal-Wallis test: $\chi^2(3, N = 65) = 38.877$, $p < .001$; Mann-Whitney U test for Group NR1 – NR2: U (N = 32) = 93.5, p = .196; in all other cases p < .01].

Figure 11 shows the mean number of blocks needed to finish different levels separately in each group. A similar pattern emerges: it seems that the task was easiest for Group WS and it was more difficult for Group WR and Group NR1 and was the most difficult for Group NR2 on all levels. On Level 1 there is significant difference between Group WS and the other groups, but there is no significant difference between Group WR, NR1 or NR2 (Kruskal-Wallis test: $\chi^2(3, N = 65) = 36.674$, $p < .001$; see U and p values from the Mann-Whitney U test for pair-wise comparisons in Table 13). This means that learning the word-pairs was the easiest when words had a meaning and were semantically related, which is not surprising.

On Level 2 there was no significant difference between Group WS and WR, WR and NR1, and between Group NR1 and NR2[15], but the difference was significant between Group WS and NR1, WS and NR2, and WR and NR2 (Kruskal-Wallis test:

[15] These results are not affected by the fact that we used 20 blocks as the measure of performance of unsuccessful participants on Level 2; these differences wouldn't have been significant even if participants continued their training for more than 20 blocks.

$\chi^2(3, N = 65) = 17.384$, $p = 0.001$; for the results of the Mann-Whitney *U* test see Table 13). This means that learning the grammar was not facilitated by the semantic relationship between words alone (Vocabulary WS vs. WR), by using words instead of familiarly sounding non-words (Vocabulary WR vs. NR1), or by the phonetic familiarity of non-words (Vocabulary NR1 vs. NR2). In other words if two vocabularies were different along one criterion only (see Table 12), it did not make the task of learning CER significantly easier. However, difference along two or three criteria significantly decreased the number of training blocks participants needed to learn the rule.

Table 13. Results of the Mann-Whitney U-test on the pair-wise analysis of the performance of groups on different levels of the task. Significant differences are emphasized by bold numbers.

Groups		Level 1	Level 2	Level 3	All Levels
WS and WR	*U*	6.000	93.000	126.500	17.000
	p	**0.000**	0.135	0.762	**0.000**
	N	33	33	33	33
WS and NR1	*U*	3.000	81.000	62.500	11.000
	p	**0.000**	**0.030**	**0.014**	**0.000**
	N	34	34	32	34
WS and NR2	*U*	6.000	41.500	41.000	9.500
	p	**0.000**	**0.000**	**0.018**	**0.000**
	N	34	34	28	34
WR and NR1	*U*	105.000	88.000	61.500	55.000
	p	0.572	0.216	0.057	**0.009**
	N	31	31	29	31
WR and NR2	*U*	110.500	56.000	39.000	41.000
	p	0.711	**0.011**	**0.048**	**0.001**
	N	31	31	25	31
NR1 and NR2	*U*	122.500	91.500	65.000	93.500
	p	0.838	0.171	0.796	0.196
	N	32	32	24	32

Pair-wise comparison of groups on Level 3 (Kruskal-Wallis test: $\chi^2(3, N = 65) = 12.296$, $p = 0.006$; for the results of the Mann-Whitney *U* test see Table 13) yielded similar results as on Level 2. This means that the same factors that helped recognizing the rule also helped generalizing and applying it to longer sentences.

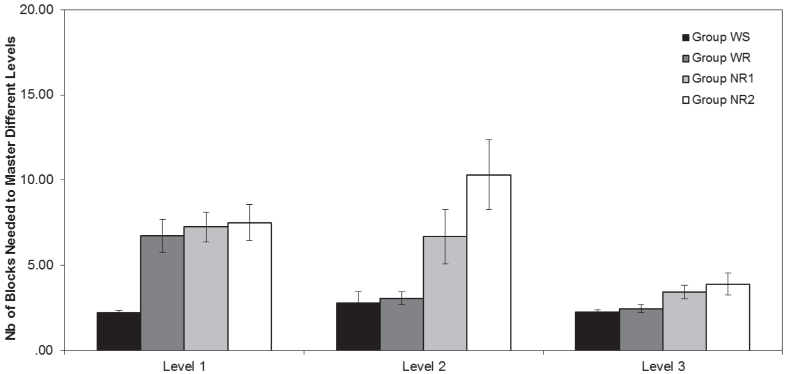


Figure 11. The mean (+-SE) number of blocks needed to master Level 1, 2 and 3 in the four groups of participants. On Level 1 there was significant difference between Group WS and all the other groups. On Level 2 and Level 3 the difference was not significant between Group WS and WR, Group WR and NR1 and Group NR1 and NR2 (those groups whose performance is represented by columns next to each other), but all other pair-wise comparisons showed significant differences.

An additional analysis was performed to compare the "words" versus "non-words" condition, which divided the participants along one dimension (whether the vocabulary was composed of natural words) into two almost equal groups. The "words" condition included participants form group WS and WR and the "non-word" condition included participants from group NR1 and NR2. The difference between the two groups was extremely significant on all levels: Level 1: U (N = 65) = 224.5, p < .001; Level 2: U (N = 65) = 266.5, p < .001; Level 3: U (N = 57) = 204.0, p = .001.

6.5 Conclusions

The present study investigated the effects of different vocabularies on the speed of learning CER in an AGL task. Sentences composed from these vocabularies represented four different degrees of diversion from natural language according to three factors: familiarity of sounds, meaning of words and semantic relationship between words (see Table 12 and Table 13). We predicted that participants trained with more realistic vocabularies would learn faster than participants trained with vocabularies less similar to natural language.

The most similar to natural language is Vocabulary WS (words semantically paired); however there still are a lot of differences. For example, in Vocabulary WS both classes of words are nouns, whereas in natural language members of word-pairs in centre-embedded sentences are from different grammatical categories (e.g.: In the sentence "*The rat that the cat chased squeaked" cat-chased* and *rat-squeaked* form word-pairs). Moreover, in natural language words can have more than one pair from a different class (e.g., *cat-ate* would also be a valid word-pair in the above mentioned sentence). Also, sentences composed from Vocabulary WS lack the dependencies between phrases present in natural sentences. On the other hand, these sentences are closer to natural language than those in other experiments in the second generation of AGL studies, because the within-phrase dependencies connecting word-pairs are semantic in nature as opposed to phonological cues used elsewhere.

Stimuli were staged according to the length of the sentences. On Level 1 of training, two-word-long sentences were presented, which required associative learning of word-pairs. It can be thought of as a simple memory task. Our analysis showed that pre-existing semantic relationships between words helped establishing these associations, but none of the other factors present in the vocabularies made a difference.

Level 2 (four-word-long sentences) involves learning or recognizing the centre-embedded structure of sentences. The instructions given to participants and the feedback presumably encouraged active rule searching as opposed to passive, incidental learning. From their written formulation of the rule it turned out that half of the unsuccessful participants were indeed involved in active rule searching because they mentioned different incorrect rules that they investigated. On this level, there was no significant difference between the learning speeds of participants who were trained with vocabularies differing in only one criterion (Table 12). Although there was significant difference between all other groups, which means that the combined effect of these criteria can help learning the grammar. The comparison of the "words" versus "non-words" condition, which yielded highly significant differences between these two groups on all levels, also supports this hypothesis.

Level 3 tests generalization of the rule to six-word-long sentences. Participants rarely scored under 80% in these blocks, which means that generalization was relatively easy. We assume that the differences between groups arose mainly from differences in the difficulty of applying the rule to the sentences. For this, remembering the first half of the sentence is required to be able to match the words with the second half of the sentence. Analysing the number of training blocks needed to pass the criterion on this level gained similar results as on Level 2, which means that the same factors that help recognizing the rule also help generalizing and applying it.

Vocabulary NR2 was very similar to the vocabulary in Bahlmann et al.' study (Bahlmann, et al., 2008). German participants in their study needed 9.47 blocks on average to finish all three levels, while Hungarian participants in our study needed 20.25 blocks. The reason for this difference could be that the vocabulary sounded more familiar to German participants than to Hungarian participants. Some participants in our study reported that they tried to associate non-words with similarly sounding words and thus giving meaning to non-words. This strategy to remember the vocabulary is obviously easier when words are phonetically closer to the participants' mother tongue.

With this in mind, we can consider the three factors listed in Table 12 as different forms or levels of semanticity: semantic relationship between words, semantic content of words (real words vs. non-words) and the ease with which non-words can be associated with some meaning. This means that semanticity of vocabularies in general influences the speed of learning.

Human participants apparently have difficulties in recognizing CER in AGL tasks: 25% of our participants did not learn the rule after 400 training sentences, when these sentences were composed of non-words with associative relationship between them and our experiment is not the only one where learning was unsuccessful (Perruchet & Rey, 2005a; de Vries, et al., 2008). This is quite contrary to the theory that CER, as an example of context-free grammar (Corballis, 2007b), is a crucial component of all human languages (W. Tecumseh Fitch & Hauser, 2004). This contradiction could be explained if there were different mechanisms at work when parsing CER in natural and in artificial languages. Maybe the factors present in natural language but absent from AGL tasks (such as the semantic content of sentences and the presence and nature of between and within-phrase dependencies) trigger those mechanisms that are responsible for parsing CER in language. This would mean that it is impossible to test the recursive component of language independently of language itself (or at least some features of language, such as semanticity). Another possibility is that CER is not parsed recursively – because multiple embeddings are practically absent from natural language, it is indeed not necessary.

In sum, the type of vocabulary does have an effect on the learnability of centre-embedded recursion. The more similar the vocabulary is to that of natural language the easier it is to learn the rule. This makes the comparison of different studies that use different vocabularies and participants with different mother tongues problematic. It also raises the question whether AGL tasks with artificial vocabularies are suitable to study the learning and processing of linguistic centre-embedded recursion. A next step in AGL experiments would be to add dependency between phrases, which in turn would make artificial sentences more similar to natural language.

7 Future directions

7.1 Concluding remarks

Since the experiment of Fitch and Hauser (2004) recursion research developed a lot, but it is still in its infancy. It is established now that while artificial sentences conforming to counting recursion *can* be parsed recursively, it is quite improbable that a learner would do that. Most often, learners of an artificial language (including models, humans and animals) will learn only regularities that are necessary for parsing the sentences and these will not always coincide with the rules that the composer of the sentences had in mind. After recognizing this, within-phrase dependencies (word-pairs) were slowly introduced to AGL experiments which insured that learners learn the centre-embedded structure of sentences.

However, learning CER does not mean that online processing involves recursion too. It is possible that while the deduced abstract rule involves centre-embedding, online processing is sequential, especially, since between-phrase dependencies are still missing from AGL experiments which would ensure that each embedding increases the complexity of sentences.

The classic training paradigm of AGL experiments included habituation by passive listening to grammatically correct stimuli of various lengths which is most probable to invoke implicit, incidental learning. This was replaced by a training paradigm that involves starting small with staged input and encourages explicit learning and active rule searching by providing feedback. The reason for this change was rather practical: participants did not learn CER by simple habituation within a reasonable time with the former method. The most common critique of recent AGL experiments is that the new training paradigm does not model correctly the learning mechanisms of infants which were in the focus of most classic AGL experiments. The problem is not with starting small, since Motherese also involves simplified grammar and vocabulary, but with providing feedback. Infants very rarely receive feedback from their environment regarding the grammaticality of their utterances. However, modelling language acquisition on adults is problematic anyway, since adults are more prone to using analytic problem solving and active rule searching than infants even without feedback.

7.2 Future directions

Recursion research faces several challenges which point to four interesting directions for future research:

1. Introducing between-phrase dependencies

To fully model recursion in natural sentences, between-phrase dependencies should be introduced to artificial sentences. In natural language semantics and morphology establishes these links, which are both absent from artificial languages. It is hard to see how to introduce these dependencies to AGL without at least semantics. However, introducing semantics makes comparative research impossible which should be in the focus of recursion research.

2. Testing online processing

As I stated above and in the introduction, learning the centre-embedded rule does not ensure that online processing is recursive. Some of the participants in our last experiment described how they processed the sentences during testing (when they already knew the rule): they continuously repeated the first half of the sentence to keep it in working memory until a pair turned up and then deleted the last word from the sequence held in memory and then continued repeating the remaining words. This is not what a stack would do, where after deleting the third word the second would automatically turn up. It does not involve self-reference either which is the main characteristic of recursion.

A human-specific way of examining online processing could use eye-tracking techniques during reading either natural or artificial sentences. Another way is to measure the processing demands that would vary according to the different candidate mechanisms. This could be achieved by measuring performance on a parallel task. The problem is that it is not possible to list all the possible mechanisms; nevertheless, testing just the most plausible one would bring us closer to the answer. At the moment, brain imaging cannot prove recursive online processing, since we do not know which brain areas are responsible for these computations (we do not even know whether CER in natural language is parsed recursively).

3. Comparative research

If there was a test for online recursive processing, the following steps should be taken to gain interesting results:

- Test online processing in humans in natural language. If natural language is parsed recursively, then
- Establish an artificial language and training method that ensures that humans process these artificial string recursively. Use the test for online processing to prove this.
- Use the same method to train and test animals. For this the artificial language and the training method should not involve anything that animals supposedly cannot process (e.g., real words).

Instead of these logical steps, early research assumed that since humans are able to parse natural sentences with recursion they must be able to do so with artificial sentences. Human parsing was not even in the focus of research, because it was assumed that it is recursive. Then, animals were tested with the same methods to see how their abilities compare to that of humans. This would have been interesting provided that human recursive parsing had had been proved before.

4. Recursion during language development

The strands of research outlined above are not interested in how recursion develops. The only question is whether recursive processing exists in human adults and nonhuman animals. Another interesting question to ask is how recursion is learnt during human development. This should involve more realistic training methods and probably infants as participants. In this strand the focus is on learning. Other strands, for example brain imaging research that tries to identify brain areas responsible for recursion, could free themselves from these restrictions in training methods and either explicitly tell participants the rule before testing or train them in the quickest possible way. If there is a module in the brain responsible for recursion, it is plausible to assume that it will do its jobs independently from how the task was introduced.

8 Acknowledgements

I would like to thank my supervisor, Eörs Szathmáry for his support, inspiration and professional advice.

I would like to thank my family and fiancée for their encouragement and understanding.

I would like to thank all my co-authors for their work.

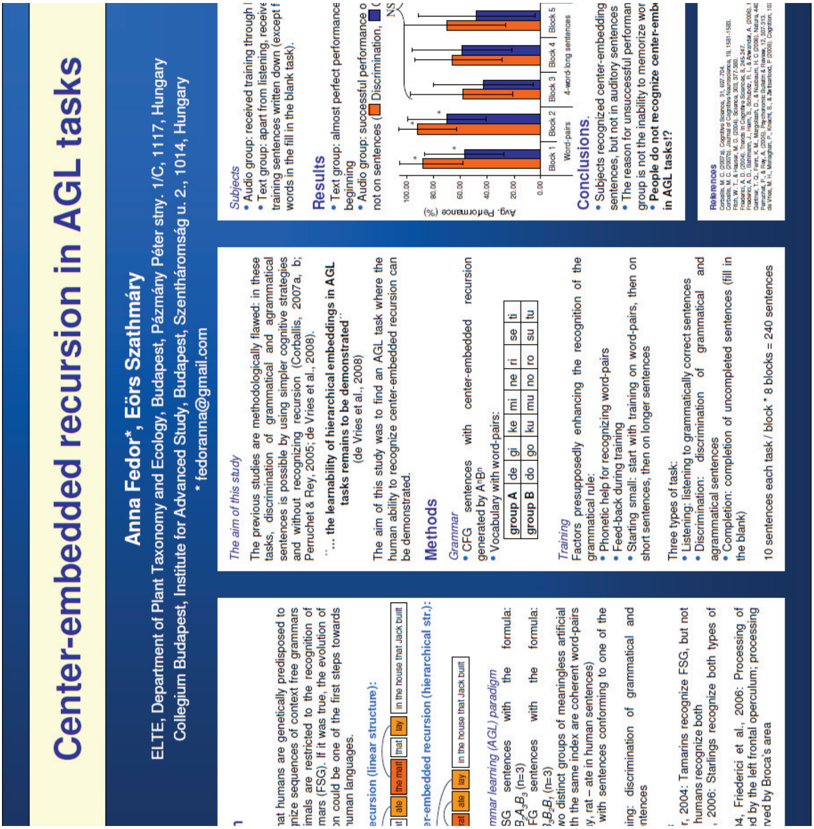
Specific acknowledgements go to: Vera Vasas for suggestions on statistics and István Zachar for blind scoring in Chapter 6. Thanks for the useful suggestions of Máté Lengyel, Chrisantha T. Fernando, Ferenc Huszár and Zoltán Szatmáry on Chapter 3.

I would like to thank all the participants who volunteered in the experiment.

Financial funds are gratefully acknowledged from the National Office for Research and Technology (NAP 2005/KCKHA005), OTKA (National Scientific Research Fund, NK73047) and European Union Framework 6 (Grant IST-FET 1940).

9 Appendix

9.1 Recursion conference poster



9.2 Score sheets for participants in the Semantics and Working memory experiment

9.2.1 Reading group

I. BLOKK

A) Szöveghallgatás (10 mondat)

Ri ro. Ke ku. Ne no. De do. Ti tu. Ri ro. Gi go. Se su. Ti tu. Ri ro.

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

1.	De ro.	igen / nem
2.	Ke ku.	igen / nem
3.	Se su.	igen / nem
4.	De ku.	igen / nem
5.	Ri ro.	igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

6. Ke ___.
7. Gi ___.
8. Ke ___.
9. Ne ___.
10. De ___.

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

11.	Ne no.	igen / nem
12.	Ne no.	igen / nem
13.	Ke su.	igen / nem
14.	Ri ro.	igen / nem
15.	De ro.	igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

16. Ri ___.
17. Ne ___.
18. Ne ___.
19. De ___.
20. Ri ___.

II. BLOKK

A) Szöveghallgatás (10 mondat)

Ri ro. Ti tu. Ke ku. Mi mu. Mi mu. Ne no. Se su. De do. Gi go. Ti tu.

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

21. De ro. igen / nem
22. Ri su. igen / nem
23. Se su. igen / nem
24. Ne su. igen / nem
25. Ke ku. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

26. Mi ___.
27. Ri ___.
28. Ti ___.
29. Ne ___.
30. Gi ___.

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

31. Mi ku. igen / nem
32. Ri ro. igen / nem
33. Gi tu. igen / nem
34. Ke do. igen / nem
35. Ri ro. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

36. Ke ___.
37. Ne ___.
38. Mi ___.
39. Mi ___.
40. Ri ___.

III. BLOKK

A) Szöveghallgatás (10 mondat)

De ke ku do. De ne no do. Gi ne no go. Ke ti tu ku. Ti ke ku tu. Gi se su go. Ti ne no tu. De ri ro do. Gi ne no go. De ke ku do.

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

41. Ti se tu su. igen / nem
42. De ke ku do. igen / nem
43. De mi mu do. igen / nem
44. Ne se no su. igen / nem
45. De se su do. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

46. De se ___ do.
47. Gi ne ___ go.
48. Ti ri ro ___ .
49. Ke gi go ___ .
50. Se ri ro ___ .

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

51. Ri de ro do. igen / nem
52. Ke se su ku. igen / nem
53. Ri de ro do. igen / nem
54. Ne se no su. igen / nem
55. De gi go do. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

56. Mi de ___ mu.
57. Ti mi ___ tu.
58. Ri mi mu ___ .
59. Ne ke ku ___ .
60. Ne ri ro ___ .

IV. BLOKK

A) Szöveghallgatás (10 mondat)

Ke mi mu ku. Ti de do tu. Ke gi go ku. Gi ri ro go. Ti gi go tu. Gi mi mu go. Mi gi go mu. Ke gi go ku. Se de do su. Gi ri ro go.

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

61.	De ne no do.	igen / nem
62.	Mi ri mu ro.	igen / nem
63.	Gi ke go ku.	igen / nem
64.	Se ne no su.	igen / nem
65.	Se ti su tu.	igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

66. Se de ___ su.
67. Se ri ___ su.
68. Mi ri ro ___.
69. Se gi go ___.
70. Ne se su ___.

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

71.	Ri ti ro tu.	igen / nem
72.	Ri mi mu ro.	igen / nem
73.	Mi gi mu go.	igen / nem
74.	Gi se su go.	igen / nem
75.	Mi ti mu tu.	igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

76. Ne mi ___ no.
77. Ne se ___ no.
78. Mi ti tu ___.
79. Ti gi go ___.
80. De ti tu ___.

V. BLOKK

A) Szöveghallgatás (10 mondat)

Ne de do no. Ke ti tu ku. Ri de do ro. Gi mi mu go. Se mi mu su. Ti se su tu. Ti mi mu tu. Ri ti tu ro. Mi ne no mu. De se su do.

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

81. Ne ke ku no. igen / nem
82. Ri de ro do. igen / nem
83. De se su do. igen / nem
84. Ne se su no. igen / nem
85. Ne se no su. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

86. Ti se su ___.
87. Se gi go ___.
88. Gi mi ___ go.
89. Ne de ___ no.
90. Ri ti ___ ro.

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

91. Mi ti tu mu. igen / nem
92. Mi se mu su. igen / nem
93. Ri gi go ro. igen / nem
94. Gi ke ku go. igen / nem
95. Ke ri ro ku. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

96. Ri mi mu ___.
97. Ti ke ku ___.
98. Ri ne ___ ro.
99. Ke ti ___ ku.
100. Ri mi ___ ro.

VI. BLOKK

A) Szöveghallgatás (10 mondat)

De ne ri ro no do. Ti gi mi mu go tu. Ne mi gi go mu no. Ke se ne no su ku. Se gi de do go su. De ne ke ku no do. Ti ke mi mu ku tu. Ke ri gi go ro ku. Se ke ri ro ku su. Ti ne mi mu no tu.

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

101.	Ke de mi mu do ku.	igen / nem
102.	Mi se ti su tu mu.	igen / nem
103.	De mi gi mu go do.	igen / nem
104.	Mi ti se tu su mu.	igen / nem
105.	Ke ti mi mu tu ku.	igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

106. Ne ri gi ___ ro no.
107. Ti ke ri ro ___ tu.
108. Gi ti ne no tu ___.
109. Mi ne se ___ no mu.
110. Gi ti mi mu ___ go.

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

111.	De ti ke tu ku do.	igen / nem
112.	Ri gi mi mu go ro.	igen / nem
113.	Gi se de do su go.	igen / nem
114.	Ne mi ti mu tu no.	igen / nem
115.	Ri ke gi go ku ro.	igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

116. Se de ti ___ do su.
117. Gi ri se su ___ go.
118. Ti de ke ku do ___.
119. Ti mi ne ___ mu tu.
120. Se ke mi mu ___ su.

VII. BLOKK

A) Szöveghallgatás (10 mondat)

Ke ne ri ro no ku. De ne se su no do. Ri de mi mu do ro. Gi mi ri ro mu go. Ke ne ti tu no ku. De se ne no su do. Mi ri ti tu ro mu. Se ke ne no ku su. De mi ne no mu do. Se mi gi go mu su.

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

121.	Ne de mi mu do no.	igen / nem
122.	Ke gi se su go ku.	igen / nem
123.	Ri de ti do tu ro.	igen / nem
124.	Gi ke ti tu ku go.	igen / nem
125.	Ne gi ri go ro no.	igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

126. Se ti ne no tu ___.
127. Ne ke se ___ ku no.
128. Mi gi ne no ___ mu.
129. Ti mi ke ku mu ___.
130. Ri ti de ___ tu ro.

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

131.	De ke se su ku do.	igen / nem
132.	Ri ne de no do ro.	igen / nem
133.	Ti de ne do no tu.	igen / nem
134.	Mi ne ti tu no mu.	igen / nem
135.	Ti se ri ro su tu.	igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

136. Mi ne de do no ___.
137. Se mi ri ___ mu su.
138. Ke mi ri ro ___ ku.
139. Ri gi se su go ___.
140. Ri mi ke ___ mu ro.

VIII. BLOKK

A) Szöveghallgatás (10 mondat)

Gi ne ri ro no go. Ne se ke ku su no. Ke mi se su mu ku. Gi de ke ku do go. Ti mi ri ro mu tu. Ri se ke ku su ro. De ti gi go tu do. De ke gi go ku do. Ti ri mi mu ro tu. Ri ne gi go no ro.

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

141.	Ri ti ke tu ku ro.	igen / nem
142.	Ke ti ne no tu ku.	igen / nem
143.	Ne ri de do ro no.	igen / nem
144.	Se ne ri ro no su.	igen / nem
145.	Ri ke de ku do ro.	igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

146. Gi de se su ___ go.
147. Gi de mi mu do ___.
148. Ti gi ne ___ go tu.
149. Mi ke de do ___ mu.
150. Gi ti ke ku tu ___.

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

151.	Mi se ke ku su mu.	igen / nem
152.	Gi ne ke ku no go.	igen / nem
153.	Se ne mi no mu su.	igen / nem
154.	De mi se mu su do.	igen / nem
155.	Se ri gi ro go su.	igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

156. Ri de ke ku ___ ro.
157. Ne se de do su ___.
158. Se de ne ___ do su.
159. Ke de ne no ___ ku.
160. Ke mi ti tu mu ___.

Szabálykeresési nyelvi kísérlet

ELTE Apáczai Csere János Gyakorlógimnázium, 12. A osztály, 2009. március 24.

A csoport

Szerinted mi volt a szabály? Fogalmazd meg egy-két mondatban!

Diktálás: sorban írd le a hallott szavakat!

Kérlek töltsd ki az alábbi kérdőívet! A megfelelő válasz betűjelét karikázd be, vagy írd be a saját válaszod!

1) Kor: …… év

2) Nem:
   a) Férfi
   b) Nő

3) Melyik kezeddel írsz?
   a) Jobb
   b) Bal

4) Mi az anyanyelved?
   a) Magyar
   b) Más: …………………………………………………………………………

5) Ha kétnyelvű vagy, vagy anyanyelvi szinten beszélsz legalább két nyelven, azt írd ide! Add meg, hogy mely nyelvek ezek, illetve mióta tanulod/ismered őket!

   a) ……………………………………………………………………………………

   b) ……………………………………………………………………………………

6) Van valamilyen krónikus betegséged? Szedsz rendszeresen valamilyen gyógyszert?

   a) Igen: …………………………………………………………………………
   b) Nem

7) Tudsz róla, hogy esetleg diszlexiás, vagy diszgráfiás vagy?
   a) Igen, diszlexiás és diszgráfiás is vagyok.
   b) Igen, diszlexiás vagyok.

c) Igen, diszgráfiás vagyok.
d) Nem, egyik sem vagyok.

8) Jól látható számodra ez a szöveg? Könnyen el tudod olvasni?
a) Igen
b) Nem

9) Jó a hallásod? (Kérlek majd a kísérlet közben is jelentkezz, ha nem hallod jól a szöveget!)
a) Igen
b) Nem

10) Beleegyezel, hogy a kísérlet során megadott adataidat és válaszaidat használjam a tanulmányomhoz (természetesen névtelenül)?
a) Igen
b) Nem

9.2.2 Listening group

I. BLOKK

A) Szöveghallgatás (10 mondat)

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

161. igen / nem
162. igen / nem
163. igen / nem
164. igen / nem
165. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

166. ______
167. ______
168. ______
169. ______
170. ______

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

171. igen / nem
172. igen / nem
173. igen / nem
174. igen / nem
175. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

176. ______
177. ______
178. ______
179. ______
180. ______

II. BLOKK

A) Szöveghallgatás (10 mondat)

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

181. igen / nem
182. igen / nem
183. igen / nem
184. igen / nem
185. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

186. ______
187. ______
188. ______
189. ______
190. ______

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

191. igen / nem
192. igen / nem
193. igen / nem
194. igen / nem
195. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

196. ______
197. ______
198. ______
199. ______
200. ______

III. BLOKK

A) Szöveghallgatás (10 mondat)

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

201. igen / nem
202. igen / nem
203. igen / nem
204. igen / nem
205. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

206. ______
207. ______
208. ______
209. ______
210. ______

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

211. igen / nem
212. igen / nem
213. igen / nem
214. igen / nem
215. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

216. ______
217. ______
218. ______
219. ______
220. ______

IV. BLOKK

A) Szöveghallgatás (10 mondat)

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

221. igen / nem
222. igen / nem
223. igen / nem
224. igen / nem
225. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

226. ______
227. ______
228. ______
229. ______
230. ______

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

231. igen / nem
232. igen / nem
233. igen / nem
234. igen / nem
235. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

236. ______
237. ______
238. ______
239. ______
240. ______

V. BLOKK

A) Szöveghallgatás (10 mondat)

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

241. igen / nem
242. igen / nem
243. igen / nem
244. igen / nem
245. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

246. ______
247. ______
248. ______
249. ______
250. ______

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

251. igen / nem
252. igen / nem
253. igen / nem
254. igen / nem
255. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

256. ______
257. ______
258. ______
259. ______
260. ______

VI. BLOKK

A) Szöveghallgatás (10 mondat)

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

261. igen / nem
262. igen / nem
263. igen / nem
264. igen / nem
265. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

266. ______
267. ______
268. ______
269. ______
270. ______

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

271. igen / nem
272. igen / nem
273. igen / nem
274. igen / nem
275. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

276. ______
277. ______
278. ______
279. ______
280. ______

VII. BLOKK

A) Szöveghallgatás (10 mondat)

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

281. igen / nem
282. igen / nem
283. igen / nem
284. igen / nem
285. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

286. ______
287. ______
288. ______
289. ______
290. ______

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

291. igen / nem
292. igen / nem
293. igen / nem
294. igen / nem
295. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

296. ______
297. ______
298. ______
299. ______
300. ______

VIII. BLOKK

A) Szöveghallgatás (10 mondat)

B) Igen/nem feladat megoldással. Megfelelnek-e a következő mondatok a szabálynak?

301. igen / nem
302. igen / nem
303. igen / nem
304. igen / nem
305. igen / nem

C) Kitöltős feladat megoldással. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

306. ______
307. ______
308. ______
309. ______
310. ______

D) Igen/nem feladat megoldás nélkül. Megfelelnek-e a következő mondatok a szabálynak?

311. igen / nem
312. igen / nem
313. igen / nem
314. igen / nem
315. igen / nem

E) Kitöltős feladat megoldás nélkül. Mi lehet a hiányzó szó, amit egy ssz hang helyettesít a mondatban?

316. ______
317. ______
318. ______
319. ______
320. ______

Szabálykeresési nyelvi kísérlet

ELTE Apáczai Csere János Gyakorlógimnázium, 12. A osztály, 2009. március 24.

A csoport

Szerinted mi volt a szabály? Fogalmazd meg egy-két mondatban!

Diktálás: sorban írd le a hallott szavakat!

Kérlek töltsd ki az alábbi kérdőívet! A megfelelő válasz betűjelét karikázd be, vagy írd be a saját válaszod!

11) Kor: …… év

12) Nem:
   a) Férfi
   b) Nő

13) Melyik kezeddel írsz?
   a) Jobb
   b) Bal

14) Mi az anyanyelved?
   a) Magyar
   b) Más: ……………………………………………………………………

15) Ha kétnyelvű vagy, vagy anyanyelvi szinten beszélsz legalább két nyelven, azt írd ide! Add meg, hogy mely nyelvek ezek, illetve mióta tanulod/ismered őket!

   a) ………………………………………………………………………………

   b) ………………………………………………………………………………

16) Van valamilyen krónikus betegséged? Szedsz rendszeresen valamilyen gyógyszert?

   a) Igen: ……………………………………………………………………
   b) Nem

17) Tudsz róla, hogy esetleg diszlexiás, vagy diszgráfiás vagy?
   a) Igen, diszlexiás és diszgráfiás is vagyok.
   b) Igen, diszlexiás vagyok.

c) Igen, diszgráfiás vagyok.

d) Nem, egyik sem vagyok.

18) Jól látható számodra ez a szöveg? Könnyen el tudod olvasni?

a) Igen

b) Nem

19) Jó a hallásod? (Kérlek majd a kísérlet közben is jelentkezz, ha nem hallod jól a szöveget!)

a) Igen

b) Nem

20) Beleegyezel, hogy a kísérlet során megadott adataidat és válaszaidat használjam a tanulmányomhoz (természetesen névtelenül)?

a) Igen

b) Nem

9.3 Supplementary material: Training and testing sentences for Vocabulary WS

This supplementary material includes the entire pool of grammatical and ungrammatical sentences for Vocabulary WS from which training and test sentences were randomly chosen for each participant in Group WS. For a description of how the sentences were composed, see the paper. Sentences were shown on a computer screen word-by-word. During presentation the first word of sentences started with a capital letter and sentences were closed by a full-stop.

Grammatical sentences
eb ól
én te
év ősz
fű fa
íny íz
kő út
eb én te ól
eb év ősz ól
eb fű fa ól
eb íny íz ól
eb kő út ól
én év ősz te
én fű fa te
én íny íz te
én kő út te
év fű fa ősz
év íny íz ősz
év kő út ősz
fű íny íz fa
fű kő út fa
íny kő út íz
én eb ól te
év eb ól ősz
fű eb ól fa
íny eb ól íz
kő eb ól út
év én te ősz
fű én te fa
íny én te íz
kő én te út
fű év ősz fa
íny év ősz íz
kő év ősz út
íny fű fa íz
kő fű fa út
kő íny íz út
eb én év ősz te ól
eb én fű fa te ól
eb én íny íz te ól
eb én kő út te ól
eb év fű fa ősz ól
eb év íny íz ősz ól
eb év kő út ősz ól
eb fű íny íz fa ól
eb fű kő út fa ól
eb íny kő út íz ól
én év fű fa ősz te
én év íny íz ősz te
én év kő út ősz te
én fű íny íz fa te
én fű kő út fa te
én íny kő út íz te
év fű íny íz fa ősz
év fű kő út fa ősz
év íny kő út íz ősz
fű íny kő út íz fa
én eb év ősz ól te
én eb fű fa ól te
én eb íny íz ól te
én eb kő út ól te
év eb fű fa ól ősz
év eb íny íz ól ősz
év eb kő út ól ősz
fű eb íny íz ól fa
fű eb kő út ól fa
íny eb kő út ól íz
év én fű fa te ősz
év én íny íz te ősz
év én kő út te ősz
fű én íny íz te fa
fű én kő út te fa
íny én kő út te íz
fű év íny íz ősz fa
fű év kő út ősz fa
íny év kő út ősz íz
íny fű kő út fa íz
év eb én te ól ősz
fű eb én te ól fa
íny eb én te ól íz
kő eb én te ól út
fű eb év ősz ól fa
íny eb év ősz ól íz
kő eb év ősz ól út
íny eb fű fa ól íz
kő eb fű fa ól út
kő eb íny íz ól út
fű én év ősz te fa
íny én év ősz te íz
kő én év ősz te út
íny én fű fa te íz
kő én fű fa te út
kő én íny íz te út
íny év fű fa ősz íz
kő év fű fa ősz út
kő év íny íz ősz út
kő fű íny íz fa út
eb év én te ősz ól
eb fű én te fa ól
eb íny én te íz ól
eb kő én te út ól
eb fű év ősz fa ól
eb íny év ősz íz ól
eb kő év ősz út ól
eb íny fű fa íz ól
eb kő fű fa út ól
eb kő íny íz út ól
én fű év ősz fa te
én íny év ősz íz te
én kő év ősz út te
én íny fű fa íz te
én kő fű fa út te
én kő íny íz út te
év íny fű fa íz ősz
év kő fű fa út ősz
év kő íny íz út ősz
fű kő íny íz út fa
én év eb ól ősz te
én fű eb ól fa te
én íny eb ól íz te
én kő eb ól út te
év fű eb ól fa ősz
év íny eb ól íz ősz
év kő eb ól út ősz
fű íny eb ól íz fa
fű kő eb ól út fa
íny kő eb ól út íz
év fű én te fa ősz
év íny én te íz ősz
év kő én te út ősz
fű íny én te íz fa
fű kő én te út fa
íny kő én te út íz
fű íny év ősz íz fa
fű kő év ősz út fa
íny kő év ősz út íz
íny kő fű fa út íz
év én eb ól te ősz
fű én eb ól te fa
íny én eb ól te íz
kő én eb ól te út
fű év eb ól ősz fa
íny év eb ól ősz íz
kő év eb ól ősz út
íny fű eb ól fa íz
kő fű eb ól fa út
kő íny eb ól íz út
fű év én te ősz fa
íny év én te ősz íz
kő év én te ősz út

íny fű én te fa íz
kő fű én te fa út
kő íny én te íz út
íny fű év ősz fa íz
kő fű év ősz fa út
kő íny év ősz íz út
kő íny fű fa íz út

Ungrammatical sentences where the error is at the last word

eb fa
én ól
év út
fű út
íny fa
kő ősz
eb út
én út
év ól
fű te
íny te
kő ól
eb én te íz
eb év ősz ősz
eb fű fa te
eb íny íz ősz
eb kő út íz
én év ősz út
én fű fa fa
én íny íz fa
én kő út ősz
év fű fa ól
év íny íz te
év kő út ól
fű íny íz íz
fű kő út ól
íny kő út te
én eb ól ól
év eb ól íz
fű eb ól te
íny eb ól ól
kő eb ól ősz
év én te íz
fű én te ősz
íny én te út
kő én te íz
fű év ősz te
íny év ősz út
kő év ősz íz
íny fű fa ősz
kő fű fa fa
kő íny íz ősz
eb én év ősz te fa
eb én fű fa te út
eb én íny íz te ősz
eb én kő út te te
eb év fű fa ősz te
eb év íny íz ősz ősz
eb év kő út ősz út
eb fű íny íz fa íz
eb fű kő út fa te
eb íny kő út íz íz
én év fű fa ősz út
én év íny íz ősz fa
én év kő út ősz út
én fű íny íz fa fa
én fű kő út fa ősz
én íny kő út íz ól
év fű íny íz fa út
év fű kő út fa fa
év íny kő út íz ól
fű íny kő út íz te
én eb év ősz ól ősz
én eb fű fa ól ősz
én eb íny íz ól fa
én eb kő út ól fa
év eb fű fa ól te
év eb íny íz ól fa
év eb kő út ól ól
fű eb íny íz ól ól
fű eb kő út ól út
íny eb kő út ól út
év én fű fa te fa
év én íny íz te íz
év én kő út te íz
fű én íny íz te ősz
fű én kő út te íz
íny én kő út te ősz
fű év íny íz ősz út
fű év kő út ősz ősz
íny év kő út ősz út
íny fű kő út fa út
év eb én te ól fa
fű eb én te ól ősz
íny eb én te ól ősz
kő eb én te ól ősz
fű eb év ősz ól út
íny eb év ősz ól te
kő eb év ősz ól ősz
íny eb fű fa ól fa
kő eb fű fa ól ősz
kő eb íny íz ól ól
fű én év ősz te út
íny én év ősz te út
kő én év ősz te ősz
íny én fű fa te fa
kő én fű fa te fa
kő én íny íz te ól
íny év fű fa ősz ól
kő év fű fa ősz fa
kő év íny íz ősz ősz
kő fű íny íz fa íz
eb év én te ősz ősz
eb fű én te fa íz
eb íny én te íz te
eb kő én te út ősz
eb fű év ősz fa ősz
eb íny év ősz íz te
eb kő év ősz út íz
eb íny fű fa íz fa
eb kő fű fa út ősz
eb kő íny íz út te
én fű év ősz fa ól
én íny év ősz íz fa
én kő év ősz út ősz
én íny fű fa íz íz
én kő fű fa út út
én kő íny íz út ősz
év íny fű fa íz íz
év kő fű fa út út
év kő íny íz út fa
fű kő íny íz út út
én év eb ól ősz ősz
én fű eb ól fa íz
én íny eb ól íz ősz
én kő eb ól út ól
év fű eb ól fa ól
év íny eb ól íz út
év kő eb ól út fa
fű íny eb ól íz te
fű kő eb ól út te
íny kő eb ól út ól
év fű én te fa te
év íny én te íz íz
év kő én te út íz
fű íny én te íz ősz
fű kő én te út út
íny kő én te út út
fű íny év ősz íz íz

fű kő év ősz út ősz
íny kő év ősz út ól
íny kő fű fa út fa
év én eb ól te ól
fű én eb ól te te
íny én eb ól te ól
kő én eb ól te fa
fű év eb ól ősz te
íny év eb ól ősz te
kő év eb ól ősz íz
íny fű eb ól fa út
kő fű eb ól fa ól
kő íny eb ól íz ősz
fű év én te ősz ősz
íny év én te ősz út
kő év én te ősz ősz
íny fű én te fa út
kő fű én te fa ősz
kő íny én te íz te
íny fű év ősz fa út
kő fű év ősz fa íz
kő íny év ősz íz íz
kő íny fű fa íz ól

Ungrammatical sentences where the error is at the next to last word

eb én ősz íz
eb év ól ősz
eb fű te te
eb íny út ősz
eb kő ól íz
én év ól út
én fű út fa
én íny te fa
én kő te ősz
év fű ősz ól
év íny út te
év kő ól ól
fű íny ól íz
fű kő ősz ól
íny kő íz te
én eb fa ól
év eb fa íz
fű eb ősz te
íny eb íz ól
kő eb ősz ősz
év én út íz
fű én íz ősz
íny én ősz út
kő én fa íz
fű év fa te
íny év íz út
kő év fa íz
íny fű ősz ősz
kő fű út fa
kő íny te ősz
eb én év ősz ősz fa
eb én fű fa íz út
eb én íny íz ősz ősz
eb én kő út íz te
eb év fű fa íz te
eb év íny íz íz ősz
eb év kő út ól út
eb fű íny íz ól íz
eb fű kő út te te
eb íny kő út te íz
én év fű fa te út
én év íny íz ól fa
én év kő út te út
én fű íny íz te fa
én fű kő út ősz ősz
én íny kő út te ól
év fű íny íz ól út
év fű kő út íz fa
év íny kő út fa ól
fű íny kő út út te
én eb év ősz íz ősz
én eb fű fa te ősz
én eb íny íz ősz fa
én eb kő út út fa
év eb fű fa te te
év eb íny íz fa fa
év eb kő út ősz ól
fű eb íny íz út ól
fű eb kő út út út
íny eb kő út te út
év én fű fa íz fa
év én íny íz fa íz
év én kő út ól íz
fű én íny íz íz ősz
fű én kő út fa íz
íny én kő út ősz ősz
fű év íny íz te út
fű év kő út íz ősz
íny év kő út te út
íny fű kő út ól út
év eb én te ősz fa
fű eb én te fa ősz
íny eb én te te ősz
kő eb én te íz ősz
fű eb év ősz te út
íny eb év ősz fa te
kő eb év ősz íz ősz
íny eb fű fa te fa
kő eb fű fa fa ősz
kő eb íny íz íz ól
fű én év ősz ősz út
íny én év ősz út út
kő én év ősz út ősz
íny én fű fa fa fa
kő én fű fa ősz fa
kő én íny íz íz ól
íny év fű fa íz ól
kő év fű fa ól fa
kő év íny íz te ősz
kő fű íny íz ól íz
eb év én te fa ősz
eb fű én te ősz íz
eb íny én te te te
eb kő én te fa ősz
eb fű év ősz ősz ősz
eb íny év ősz te te
eb kő év ősz fa íz
eb íny fű fa ól fa
eb kő fű fa te ősz
eb kő íny íz ól te
én fű év ősz ősz ól
én íny év ősz út fa
én kő év ősz te ősz
én íny fű fa fa íz
én kő fű fa fa út
én kő íny íz ősz ősz
év íny fű fa út íz
év kő fű fa fa út
év kő íny íz íz fa
fű kő íny íz fa út
én év eb ól ól ősz
én fű eb ól te íz
én íny eb ól fa ősz
én kő eb ól íz ól
év fű eb ól ól ól
év íny eb ól ól út
év kő eb ól ősz fa
fű íny eb ól te te
fű kő eb ól te te
íny kő eb ól ősz ól
év fű én te út te
év íny én te ól íz
év kő én te fa íz

fű íny én te ól ősz
fű kő én te ól út
íny kő én te ősz út
fű íny év ősz út íz
fű kő év ősz ősz ősz
íny kő év ősz íz ól
íny kő fű fa íz fa
év én eb ól út ól
fű én eb ól ól te
íny én eb ól ősz ól
kő én eb ól út fa
fű év eb ól ól te
íny év eb ól ól te
kő év eb ól íz íz
íny fű eb ól út út
kő fű eb ól te ól
kő íny eb ól ól ősz
fű év én te fa ősz
íny év én te íz út
kő év én te fa ősz
íny fű én te út út
kő fű én te út ősz
kő íny én te te te
íny fű év ősz te út
kő fű év ősz ól íz
kő íny év ősz út íz
kő íny fű fa út ól

Ungrammatical sentences where the error is at the one before the next to last word

eb én év ól ősz fa
eb én fű íz íz út
eb én íny út ősz ősz
eb én kő ősz íz te
eb év fű út íz te
eb év íny ól íz ősz
eb év kő íz ól út
eb fű íny út ól íz
eb fű kő ősz te te
eb íny kő fa te íz
én év fű te te út
én év íny út ól fa
én év kő ősz te út
én fű íny ősz te fa
én fű kő fa ősz ősz
én íny kő fa te ól
év fű íny fa ól út
év fű kő fa íz fa
év íny kő íz fa ól
fű íny kő ól út te
én eb év íz íz ősz
én eb fű ól te ősz
én eb íny te ősz fa
én eb kő te út fa
év eb fű út te te
év eb íny te fa fa
év eb kő fa ősz ól
fű eb íny út út ól
fű eb kő íz út út
íny eb kő te te út
év én fű ól íz fa
év én íny fa fa íz
év én kő ősz ól íz
fű én íny ősz íz ősz
fű én kő ól fa íz
íny én kő te ősz ősz
fű év íny te te út
fű év kő íz íz ősz
íny év kő fa te út
íny fű kő te ól út
év eb én íz ősz fafű eb én íz fa ősz
íny eb én ősz te ősz
kő eb én íz íz ősz
fű eb év te te út
íny eb év íz fa te
kő eb év út íz ősz
íny eb fű íz te fa
kő eb fű út fa ősz
kő eb íny te íz ól
fű én év ól ősz út
íny én év fa út út
kő én év te út ősz
íny én fű ősz fa fa
kő én fű út ősz fa
kő én íny te íz ól
íny év fű út íz ól
kő év fű te ól fa
kő év íny ól te ősz
kő fű íny te ól íz
eb év én fa fa ősz
eb fű én íz ősz íz
eb íny én ősz te te
eb kő én íz fa ősz
eb fű év íz ősz ősz
eb íny év íz te te
eb kő év te fa íz
eb íny fű íz ól fa
eb kő fű út te ősz
eb kő íny fa ól te
én fű év fa ősz ól
én íny év te út fa
én kő év út te ősz
én íny fű ól fa íz
én kő fű út fa út
én kő íny út ősz ősz
év íny fű te út íz
év kő fű te fa út
év kő íny ól íz fa
fű kő íny fa fa út
én év eb ősz ól ősz
én fű eb út te íz
én íny eb út fa ősz
én kő eb te íz ól
év fű eb te ól ól
év íny eb te ól út
év kő eb te ősz fa
fű íny eb út te te
fű kő eb íz te te
íny kő eb te ősz ól
év fű én út út te
év íny én íz ól íz
év kő én fa fa íz
fű íny én ősz ól ősz
fű kő én íz ól út
íny kő én út ősz út
fű íny év út út íz
fű kő év út ősz ősz
íny kő év íz íz ól
íny kő fű te íz fa
év én eb íz út ól
fű én eb te ól te
íny én eb te ősz ól
kő én eb fa út fa
fű év eb fa ól te
íny év eb út ól te
kő év eb íz íz íz
íny fű eb fa út út
kő fű eb ősz te ól
kő íny eb te ól ősz
fű év én fa fa ősz
íny év én ősz íz út
kő év én ól fa ősz
íny fű én ól út út
kő fű én fa út ősz
kő íny én fa te te
íny fű év út te út
kő fű év ól ól íz

kő íny év fa út íz
kő íny fű ősz út ól

10 Summary

Recursion is a very important phenomenon not only for the origins of human language but possibly for human evolution too. There are several meanings of this term but generally, a procedure is recursive if it calls itself, and a structure is recursive if it contains a structure of the same kind. These definitions can be applied to various entities and indeed, the neuronal mechanism of recursion is held responsible for a various range of human cognitive abilities, like tool-making, theory of mind, mental time travel and counting. This thesis investigates one particular type of recursion, centre-embedded recursion (CER) which results in centre-embedded sentences, such as *The rat that the cat chased squeaked*, where the phrase (*that the cat chased*) is embedded within another phrase (*the rat squeaked*).

Here, I describe a connectionist model and three psycholinguistic experiments using the artificial grammar learning paradigm. It is known that centre-embedded sentences can be effectively parsed by a push-down automaton. Chapter 3 shows how this symbolic model can be neurally implemented with the help of gating neurons.

The aim of our first experiment (Chapter 4) was to find out whether human participants in Fitch and Hauser's (2004) experiment could learn the centre-embedded structure of artificial sentences or just learnt a simpler rule, namely counting recursion. We have found that our participants did not learn CER, although we have used similar methods.

With the second experiment (Chapter 5) we investigated the question whether learning CER can be made easier by decreasing working memory demands by providing participants with written artificial sentences. We have found that when participants received the script of the stimuli they have learned the rule almost instantly. When the sentences were auditorily presented, the task was much more difficult, in fact half of the participants could not recognize the rule.

Since it seems that human participants have difficulty learning CER in artificial sentences, although natural language contains CER, we hypothesized that this difference is caused by the lack of semantics in artificial sentences. In the third experiment (Chapter 6) we have found that the more semanticity is included in the stimuli, the faster participants learned CER. This raises the possibility that there are different mechanisms at work when parsing CER in natural and in artificial languages which in turn, would questions the suitability of the artificial grammar learning paradigm to study linguistic CER. Suggestions for the direction of future research are included.

11 Összefoglaló

A rekurzió egy fontos jelenség az emberi nyelv eredetével és általában az emberi evolúcóval kapcsolatban. Több jelentése is van ennek a kifejezésnek, de általánosságban egy procedúra akkor rekurzív, ha meghívja önmagát, egy struktúra pedig akkor rekurzív, ha tartalmaz egy hasonló típusú struktúrát. Ezek a definíciók többféle dologra is alkalmazhatók, így a rekurzió neurális mechanizmusát tartják felelősnek egy sor kognitív emberi képességért, mint például eszközkészítés, elme-elmélet, mentális időutazás és számolás. Ebben a dolgozatban egy konkrét rekurziófajtával foglalkozom, a középre beágyazott rekurzióval (KBR) melynek eredményei középre beágyazott mondatok, mint pl. *A patkány, amit a macska megkergetett, vinnyogott*, ahol az egyik frázis (*amit a macska megkergetett*) be van ágyazva a másik frázisba (*a patkány vinnyogott*).

Egy konnekcionista modellt és három pszicholingvisztikai kísérletet írok le, melyek során a mesterséges nyelvtantanulási paradigmát használtuk. Az régóta tudott, hogy a középre beágyazott mondatok hatékonyan elemezhetőek veremautomatával. A harmadik fejezetben megmutatjuk, hogy ez a szimbolikus modell hogyan implementálható neurálisan kapuzó neuronok segítségével.

Az első kísérletünk (4. fejezet) célja az volt, hogy kiderítsük, hogy Fitch és Hauser (2004)-es kísérletében a humán résztvevők megtanulhatták-e a mesterséges mondatok középre beágyazott szerkezetét, vagy csak egy egyszerűbb szabályt, a számláló rekurziót tanulták meg. A mi kísérletünkben résztvevők nem tanulták meg a KBR-t, holott hasonló módszereket használtunk.

A második kísérletünkben (5. fejezet) azt a kérdést vizsgáltuk, hogy a KBR megtanulása könnyebbé tehető-e azzal, ha a résztvevőknek odaadjuk a mesterséges mondatokat leírva, ezzel csökkentve a munkamemóriára rótt terheket. Azt találtuk, hogy amikor a résztvevők megkapták a stimulust írott formában is, majdnem rögtön megtanulták a szabályt. Amikor csak hallották a mondatokat, a feladat nehezebb volt és a résztvevők fele meg sem tudta tanulni a szabályt.

Habár a természetes nyelvekben vannak rekurzív mondatok, úgy tűnik, hogy a résztvevőknek nehézséget okoz megtanulni a KBR-t mesterséges mondatokban. Feltételeztük, hogy ezt az ellentmondást az okozza, hogy a mesterséges mondatok nem tartalmaznak szemantikus információt. A harmadik kísérletben (6. fejezet) azt találtuk, hogy minél több szemantikus információt tartalmaznak a mondatok, annál könnyebb megtanulni a KBR-t. Ez felveti annak a lehetőségét, hogy a mesterséges és természetes mondatokban jelen levő KBR elemzéséért különböző mechanizmusok felelősek, ami viszont megkérdőjelezheti a mesterséges nyelvtantanulási paradigma használhatóságát a nyelvi KBR tanulmányozásánál. A dolgozat végén javaslatok találhatók további kutatásokhoz.

12 Publications

Publications in peer-reviewed journals[16]

1. Anna Fedor, Máté Varga, Eörs Szathmáry (2012). Semantics boosts syntax in artificial grammar learning tasks with recursion. Journal of Experimental Psychology – Learning, Memory and Language, *accepted manuscript*.
   Citations: 0; Impact factor: 2.761
2. István Zachar, Anna Fedor, Eörs Szathmáry (2011). Two Different Template Replicators Coexisting in the Same Protocell: Stochastic Simulation of an Extended Chemoton Model. PLoS ONE 6(7): e21380.
   doi:10.1371/journal.pone.0021380
   Citations: 0; Impact factor: NA
3. Anna Fedor, Péter Ittzés, Eörs Szathmáry (2010). Parsing recursive sentences with a connectionist model including a neural stack and synaptic gating. Journal of Theoretical Biology, 271, 100–105.
   Citations: 0;Impact factor: 2.371
4. Anna Fedor, Vera Vasas (2009). The robustness of keystone indices in food webs. Journal of Theoretical Biology, 260, 372–378.
   Citations: 2; Impact factor: 2.371
5. Anna Fedor, Gabriella Skollár, Nóra Szerencsy, Mária Ujhelyi (2008). Object permanence tests on gibbons (*Hylobatidae*). Journal of Comparative Psychology, 122, 403-417.
   Citations: 10; Impact factor: 2.138

[16] Publications 2, 4 and 5 are not related to the present thesis.

Publications in edited books

6. Anna Fedor, Péter Ittzés and Eörs Szathmáry (2009). The biological background of syntax evolution. In: Biological foundations and origin of syntax, pp. 15-40. Eds.: Bickerton, D., and Szathmáry, E. Strüngmann Forum Report, vol. 3. Cambridge, MA: The MIT Press.
   Citations: 0
7. Anna Fedor, Jens Brauer, David Caplan, Angela D. Friederici, Balázs Gulyás, Peter Hagoort, Tatjana Nazir, Csaba Pléh and Wolf Singer (2009). What are the brain mechanisms underlying syntactic operations? In: Biological foundations and origin of syntax, pp. 299-324. Eds.: Bickerton, D., and Szathmáry, E. Strüngmann Forum Report, vol. 3. Cambridge, MA: The MIT Press.
   Citations: 2
8. Eörs Szathmáry, Zoltán Szatmáry, Péter Ittzés, Gergő Orbán, István Zachár, Ferenc Huszár, Anna Fedor, Máté Varga, Szabolcs Számadó (2007). In silico Evolutionary Developmental Neurobiology and the Origin of Natural Language. In: Emergence of Communication and Language, Springer, London. Eds: Caroline Lyon, Chrystopher Nehaniv, Angelo Cangelosi. pp. 151-187.
   Citations: 7

Publications in Hungarian journals

9. Fedor Anna, Ittzés Péter, Szathmáry Eörs (2010). A nyelv evolúciójának biológiai háttere (*Biological background of language evolution*). *Magyar Tudomány, 171*, 541-548.

13 References

Bahlmann, J., Heim, S., Schubotz, R., & Friederici, A. D. (2006). FMRI investigation of the processing of linear and hierarchical structures: An artificial grammar task. *Journal of Cognitive Neuroscience, Annual Meeting Supplement*, 126.

Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage, 42*(2), 525-534.

Bickerton, D. (1990). *Language and Species*: University Of Chicago Press.

Bickerton, D. (1995). *Language and human behaviour*: Seattle, University of Washington Press.

Bickerton, D. (2009). *Adam's Tongue: How Humans Made Language, How Language Made Humans*: New York: Hill and Wang.

Bickerton, D., & Szathmáry, E. (2011). Confrontational scavenging as a possible source for language and cooperation. *BMC Evolutionary Biology, 11*(261).

Boden, M., & Wiles, J. (2000). Context-free and context-sensitive dynamics in recurrent neural networks. *Connection Science, 12*(3-4), 197-210.

Chen, C. H., & Honavar, V. (1999). A neural-network architecture for syntax analysis. *IEEE Trans Neural Netw, 10*(1), 94-114.

Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world: Inside the mind of another species*: University of Chicago Press, Chicago.

Chomsky, N. (1956). Three models for the description of language. *Information Theory, IRE Transactions on, 2*(3), 113-124.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science: A Multidisciplinary Journal, 23*(2), 157 - 205.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*(2), 157-205.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics in perspective. In M. H. Christiansen & N. Chater (Eds.), *Connectionist Psycholinguistics* (pp. 19-76): Greenwood Publishing Group.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: capturing the empirical data. *Trends in Cognitive Sciences, 5*(2), 82-88.

Cochran, B. P., McDonald, J. L., & Parault, S. J. (1999). Too smart for their own good: The disadvantage of a superior processing capacity for adult language learners. *Journal of Memory and Language, 41*, 30-58.

Conway, C. M., Ellefson, M. R., & Christiansen, M. H. (2003). *When less is less and when less is more: Starting small with staged input.* Paper presented at the 25th Annual Conference of the Cognitive-Science-Society, Boston, MA.

Corballis, M. C. (2007a). On phrase structure and brain responses: A comment on Bahlmann, Gunter, and Friederici (2006). *Journal Of Cognitive Neuroscience, 19*(10), 1581-1583.

Corballis, M. C. (2007b). Recursion, language, and starlings. *Cognitive Science, 31*(4), 697-704.

Corballis, M. C. (2007c). The uniqueness of human recursive thinking. *American Scientist, 95*(3), 240-248.

Corballis, M. C. (2011). *The recursive mind - The Origins of Human Language, Thought and Civilization*: Princeton University Press.

Dayan, P., & Abbott, L. F. (2005). *Theoretical neuroscience: computational and mathematical modeling of neural systems.* Cambridge, MA: The MIT Press.
Deacon, T. W. (2003). Multilevel selection in a complex adaptive system: the problem of language origins. In B. H. Weber & D. J. Depew (Eds.), *Evolution and learning: The Baldwin effect reconsidered* (pp. 81-106). Cambridge, MA: MIT Press.
Edwards, D. (1973). Sensory-motor intelligence and semantic relations in early child grammar. *Cognition, 2*(4), 395-434. doi: doi:10.1016/0010-0277(73)90002-4
Eens, M. (Ed.). (1997). *Understanding the Complex Song of the European Starling: An Integrated Ethological Approach*: Academic Press, New York.
Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, And Grammatical Structure. *Machine Learning, 07*, 195-225.
Elman, J. L. (1993). Learning And Development In Neural Networks - The Importance Of Starting Small. *Cognition, 48*(1), 71-99.
Everett, D. L. (2005). Cultural constraints on grammar and cognition in Piraha - Another look at the design features of human language. *Current Anthropology, 46*(4), 621-646.
Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and brain sciences, 27*, 491–541.
Fedor, A., Ittzés, P., & Szathmáry, E. (2009). The biological background of syntax evolution. In D. Bickerton & E. Szathmáry (Eds.), *Biological foundations and origin of syntax. Strüngmann Forum Report* (Vol. 3, pp. 15-40). Cambridge, MA: The MIT Press.
Fedor, A., Ittzés, P., & Szathmáry, E. (2010). Parsing recursive sentences with a connectionist model including a neural stack and synaptic gating. *Journal of Theoretical Biology, 271*, 100–105.
Fedor, A., & Szathmáry, E. (2009). *Center-embedded recursion in AGL tasks - the role of working memory*.
Fedor, A., Varga, M., & Szathmáry, E. (2012). Semantics boosts syntax in artificial grammar learning tasks with recursion. *Journal of Experimental Psychology – Learning, Memory and Language, accepted ms*.
Fitch, W. T. (2010). Three Meanings of "Recursion": Key Distinctions for Biolinguistics. In R. Larson, V. Déprez & H. Yamakido (Eds.), *The Evolution of Human Language: Biolinguistic Perspectives* (pp. 73-90): Cambridge University Press, Cambridge, UK.
Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*(5656), 377-380.
Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*(2), 179-210.
Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America, 103*(7), 2458-2463.
Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science, 165*, 664-672.
Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*(7088), 1204-1207.
Gisiger, T., Kerszberg, M., & Changeux, J. P. (2005). Acquisition and performance of delayed-response tasks: a neural network model. *Cereb Cortex, 15*(5), 489-506.
Hauser, M. D., Chomsky, N., & Fitch, T. W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua, 120*(8), 1887-1897.
Herman, L. M., Richards, D. G., & Wolz, J. P. (1984). Comprehension of sentences by bottle-nosed dolphins. *Cognition, 16*, 129-219.
Hjelmfelt, A., Weinberger, E. D., & Ross, J. (1991). Chemical implementation of neural networks and Turing machines. *Proceedings of the National Academy of Sciences of the United States of America, 88*(24), 10983-10987.
Hopcroft, J. E., & Ullman, J. D. (1979). *Introduction to Automata Theory, Languages, and Computation.* Reading, MA: Addison-Wesley.
Hurford, J. (2007). *The Origins of Meaning*. Oxford: Oxford Univ. Press.
Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language - (Reply to Fitch, Hauser, and Chomsky). *Cognition, 97*(2), 211-225.
Jurafsky, D., & Martin, J. H. (2007). *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition.* (2 ed.): Prentice Hall.
Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics, 43*(2), 365-392.
Katz, P. S. (2003). Synaptic Gating: The Potential to Open Closed Doors. *Current biology : CB, 13*(14), R554-R556.
Kernighan, B. W., & Ritchie, D. M. (1988). *The C Programming Language* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
Lai, J., & Poletiek, F. H. (2011). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition, 118*(2), 265 - 273.
Lobina, D. J. (2011a). Recursion and the competence/performance distinction in AGL tasks. *Language and Cognitive Processes, in press*.
Lobina, D. J. (2011b). *Recursion: A computational investigation into the human representation and processing of language. Unpublished doctoral dissertation.* Universitat de Barcelona, Department of Philosophy and Universitat Rovira i Virgili, Department of Psychology.
Maynard Smith, J., & Szathmáry, E. (1995). *The Major Transitions in Evolution.* Oxford: Freeman.
Mazzoni, P., Andersen, R. A., & Jordan, M. I. (1991). A more biologically plausible learning rule for neural networks. *Proceedings of the National Academy of Sciences of the United States of America, 88*(10), 4433-4437.
Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127-162.
O'Reilly, R. C. (1996). Biologically plausible error-driven learning using local activation differences: The generalized recirculation algorithm. *Neural Comput., 8*(5), 895-938. doi: http://dx.doi.org/10.1162/neco.1996.8.5.895
O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science, 314*(5796), 91-94.
Patterson, F. (1978). Conversations with a gorilla. *National Geographic, 154*, 438-465.
Pena, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604-607.
Pepperberg, I. M. (1990). Some cognitive capacities of an African Grey parrot (Psitaccus erithacus). *Advances in the study of behavior, 19*, 357-409.
Perruchet, P., & Rey, A. (2005a). Does the Mastery of Center-Embedded Linguistic

Structures Distinguish Humans from Nonhuman Primates? *Psychonomic Bulletin & Review, 12*, 307-313.
Perruchet, P., & Rey, A. (2005b). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin \& Review, 12*(2), 307-313.
Pinker, S. (1994). *The Language Instinct.*: Penguin.
Pinker, S., & Jackendoff, R. (2005). The faculty of language: what's special about it? *Cognition, 95*(2), 201-236.
Premack, D. G., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*, 515-526.
Robinson, J. G. (1984). Syntactic Structures in the Vocalizations of Wedge-Capped Capuchin Monkeys, Cebus Olivaceus. *Behaviour, 90*, 46-78.
Rodriguez, J., Wiles, J., & Elman, J. L. (1999). A recurrent neural network that learns to count. *Connection Science, 11*, 5-40.
Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: how important is starting small? *Cognition, 72*(1), 67-109.
Rohde, D. L. T., & Plaut, D. C. (2002). Less is Less in Language Acquisition. In P. Quinlin (Ed.), *Connectionist modelling of cognitive development*: Hove, UK: Psychology Press.
Sankaranarayanan, G. (2002). Repetitive forms in Indian languages. *Language in India, 1*, online journal.
Soltoggio, A., Dürr, P., Mattiussi, C., & Floreano, D. (2007). *Evolving neuromodulatory topologies for reinforcement-like problems.* Paper presented at the IEEE Congress on Evolutionary Computation.
Sun, G. Z., Giles, C. L., Chen, H. H., & Lee, Y. C. (1998). The Neural Network Pushdown Automaton: Architecture, Dynamics and Training. In C. L. Giles & M. Gori (Eds.), *Adaptive Processing of Sequences and Data Structures* (pp. 296-345): Springer.
Szathmáry, E. (2007). The Codes of Life: The Rules of Macroevolution. In M. Barbieri (Ed.), (pp. 283-313): Springer.
Szathmáry, E., Szatmáry, Z., Ittzés, P., Orbán, G., Zachár, I., Huszár, F., . . . Számadó, S. (2007). The emergence of communication and language. In C. Lyon, C. L. Nehaniv & A. Cangelosi (Eds.): Springer.
Tettamanti, M., & Weniger, D. (2006). Broca's area: A supramodal hierarchical processor? *Cortex, 42*(4), 491-494.
de Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition, 107*(2), 763-774.
Wiles, J., Watson, J., Tonkes, B., & Deacon, T. (2005). Transient Phenomena in Learning and Evolution: Genetic Assimilation and Genetic Redistribution. *Artif. Life, 11*(1-2), 177-188. doi: http://dx.doi.org/10.1162/1064546053279026
Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour, 63*, 293-299.